

Data-Driven Parallel Scientific Computing: Multi-Fidelity Information Fusion Algorithms and Applications to Physical and Biological Systems

by

Paris Perdikaris

M.Sc., Brown University; Providence, R.I., U.S.A., 2010

M.Eng., National Technical University of Athens; Athens, Greece, 2009

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in The Division of Applied Mathematics at Brown University

PROVIDENCE, RHODE ISLAND

May 2015

© Copyright 2015 by Paris Perdikaris

This dissertation by Paris Perdikaris is accepted in its present form
by The Division of Applied Mathematics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date_____

George Em. Karniadakis, Ph.D., Advisor

Recommended to the Graduate Council

Date_____

Johannes O. Royset, Ph.D., Reader

Date_____

Daniele Venturi, Ph.D., Reader

Approved by the Graduate Council

Date_____

Peter M. Weber, Dean of the Graduate School

Vita

Paris Perdikaris was born on March 27, 1984, in Athens, Greece, and grew up in the coastal town of Paleo Faliro. He completed studies leading to a Masters of Engineering in Naval Architecture and Marine Engineering at the National Technical University of Athens, receiving the Thomaidio award for scientific publications in 2009.

As a PhD candidate in the Division of Applied Mathematics at Brown University, Paris has been advised and mentored by Prof. George Em. Karniadakis. In 2009 he received a University Fellowship for graduate studies, and completed a comprehensive two semester, 8-course training in Applied Mathematics. In addition to teaching assistantships, Paris has worked on several projects funded by major U.S agencies, all under the supervision of Prof. George Em. Karniadakis. His current work aims at setting the foundations of a new type of data-driven scientific computing, bridging traditional techniques, e.g., spectral element methods, with contemporary machine learning algorithms.

Peer-reviewed publications

1. Perdikaris, P., D. Venturi, and G. E. Karniadakis. Multi-fidelity information fusion algorithms for high dimensional systems and massive data-sets. *SIAM Journal of Scientific Computing*, (to be submitted), 2015.
2. Perdikaris, P., D. Venturi, J. O. Royset, and G. E. Karniadakis. Multi-fidelity modeling via recursive co-kriging and Gaussian Markov random fields. *Proceedings of the Royal Society A*, (under review), 2015.
3. Perdikaris, P., J. A. Insley, L. Grinberg, Y. Yu, M. E. Papka, and G. E. Karniadakis. Investigating flow-structure interactions in cerebral aneurysms. *Supercomputing*, 2014.
4. Perdikaris, P., L. Grinberg, and G. E. Karniadakis. An effective fractal-tree closure model for simulating blood flow in large arterial networks. *Annals of biomedical engineering*, 1-11, 2014.
5. Perdikaris, P. and G. E. Karniadakis. Fractional-order viscoelasticity in one-dimensional blood flow models. *Annals of biomedical engineering*, 42:1012-1023, 2014.
6. Perdikaris, P., L. Kaiktsis, and G. S. Triantafyllou. Chaos in a cylinder wake due to forcing at the Strouhal frequency. *Physics of Fluids*, 21:101705, 2009.

Preface and Acknowledgments

This thesis has been compiled over a mosaic of research projects I have been involved in over the last 6 years at Brown. It all started in the summer of 2010 when I took a leap from my naval engineering background to study the viscoelastic behavior of soft tissue. This problem introduced me to the fascinating world of modeling multi-physics phenomena in the human circulation – a biological system that is the envy of every engineer. This resulted to a series of findings that were centered around three open questions in computational hemodynamics, namely creating robust models that exhibit low sensitivity to parametrizations, addressing the closure problem for simulating blood flow in truncated arterial domains, and developing efficient techniques for model inversion and parameter calibration.

In the summer of 2014 my research activities took a very interesting and exciting twist, as I was captured by the potential application of machine learning techniques in multi-fidelity modeling, design optimization under uncertainty, data assimilation, and model inversion. I soon realized that the practical applicability of such techniques is hindered by the prohibitive computational cost of training surrogate models in the presence of large data-sets, and scaling the predictive inference schemes to high dimensional problems. This led to a dedicated two-semester effort to address these challenges and design a tractable and scalable framework that opens the path to a new direction in data-driven scientific computing.

There are many people, without whom, this work would have been infeasible.

First, I would like to sincerely thank my advisor, Prof. George Em. Karniadakis. This thesis would be impossible to be conceived, implemented, and published without his strong support. George gave me the opportunity to join his vibrant research group 6 years ago, and has been a limitless resource of inspiration, dedication, and mentorship ever since. Throughout this time, he provided an open highway for me to express myself, feed my curiosity, and develop both as a person and as a scientist. His intuition, ideas, and constructive criticism have been invaluable to the development of this thesis.

Next, I would like to genuinely thank Prof. Daniele Venturi. Daniele has been a good friend and an ideal colleague throughout my years at Brown. This thesis would have never reached its current form without his help and contributions. As we are taking different paths away from Brown, I take his work ethics and originality of thinking as examples of true scholarship, and I look forward to a productive collaboration in the years to come.

I would like to thank Prof. Johannes O. Royset for our ongoing and productive collaboration, and for taking the time to participate in my thesis committee. I also thank Dr. Leopold Grinberg for our fruitful collaboration, and for sharing his extensive experience on high performance scientific computing. I am also very grateful to Prof. George Triantafyllou, and Prof. Lambros Kaiktsis at the National Technical University of Athens that fully supported and encouraged my endeavor for advanced studies.

I owe a wholehearted thanks to all faculty members and staff at the Division of Applied Mathematics at Brown University. The achievements and deep scientific knowledge of the first have served as an ocean of inspiration, while the administrative support and efficiency of the later have simplified many every-day tasks. I would like

to especially thank Prof. Basilis Gidas for being a good friend, mentor, and resourceful explorer of Providence’s gastronomic hideouts. I also thank Prof. Constantine Dafermos for supporting my pursuit for graduate studies at Brown, and for being an example of scientific rigor and integrity. Many thanks also go to Madeline Brewster and Jean Radican for their administrative support. My gratitude also goes to the CCV, ORNL, and ALCF staff for their support and training on large-scale parallel computing, and especially to Dr. Joseph Isley for his artistic touch in scientific visualization.

I thank all my classmates and lab-mates for providing a healthy and joyful work environment. Especially, I thank the “old guard” of the 37 Manning basement: Heyrim, Mingge, Summer, Changho, Joon, Mohsen, Handy, Xiu, and Yue, for filling the space with their light and making those long nights of work a bit less lonely.

A very special thanks goes to her highness, Sophia Sa, for being the *yáng* to my *yín*, constantly challenging her limits of patience, and bringing color to the tedious days of this dissertation. I feel very grateful that I got the chance to meet and interact with many wonderful people here at Brown. I’ll probably forget some, but among all, I thank Babis for his camaraderie and thirst for adventures, Socrates for his wisdom and voice of reason, Foteini for giving me my first tour of Thayer St., Ares for his genuine friendship and unique perspectives, Vaio for letting me win at pool, Jesus for his unbearable lightness of being, Maria for her limitless amount of positive energy, Kostas and Anastasia, Alexandra, Evgenios, Dimitra, the Sa family, Mary and Christopher, and last, but not least, Tasos for being a patient roommate for over than 3 years.

A very well deserved expression of gratitude also goes to my good friends in Greece, Christos, Giannis, Mimis, and the Jinetera crew, for helping me recharge my

batteries every summer, and for constantly radiating positive vibes from 4,767 miles away.

Last but not least, I am beyond thankful to my family, Vaso, Myrto, and Angelos. If words were to exist for expressing my appreciation for their unconditional love and never-ending support, you would surely find them in this sentence. This work is dedicated to them with immense gratitude.

Contents

Vita	iv
Preface and Acknowledgments	vi
1 Introduction	1
1.1 Data-driven Scientific Computing	2
1.1.1 Part 1: Scalable multi-fidelity information fusion algorithms	3
1.1.2 Part 2: Open questions in cardiovascular mathematics	4
2 A General Framework for Multi-fidelity Information Fusion	6
2.1 Motivation	7
2.2 Multi-fidelity modeling	10
2.3 Methods	12
2.3.1 Kriging	12
2.3.2 Recursive co-kriging	14
2.3.3 Maximum likelihood estimation	16
2.4 Multi-fidelity in models and in probability space	17
2.5 Challenges to be addressed	20
2.5.1 Bottlenecks of learning from large data-sets	20
2.5.2 Bottlenecks of scaling to high dimensions	21
2.6 Demonstration	22
2.6.1 Multi-fidelity modeling of ocean waves	22
2.6.2 Data-assimilation of coastal bathymetry data	24
2.6.3 Effect of low correlation	28
3 Spectral/<i>hp</i> Element Methods for Inference with Gaussian Markov Random Fields	32
3.1 Motivation	33
3.2 Gaussian Markov random fields and the SPDE approach	33
3.2.1 The Matérn covariance model	34
3.2.2 Connection to stochastic partial differential equations (SPDE)	35
3.3 Hilbert space approximation	37
3.3.1 Weak form	37
3.3.2 Choosing an expansion basis	39

3.4	The spectral/ <i>hp</i> element method	41
3.4.1	Overview of the method	41
3.4.2	Properties of the nodal basis	43
3.4.3	Treatment of boundary conditions	44
3.4.4	Computation of the covariance and precision matrices	47
3.5	Model fitting and the kriging predictor	50
3.5.1	Maximum Likelihood Estimation with GMRFs	54
3.6	Multi-fidelity modeling via recursive co-kriging and Gaussian Markov random fields	55
3.6.1	Ordinary kriging with GMRFs	55
3.6.2	Workflow	60
3.7	Results	62
3.7.1	Helmholtz equation in 2D	62
3.7.2	A prototype problem in risk-averse design	67
3.7.3	Multi-fidelity regression of a random function	73
3.7.4	Stochastic incompressible flow	74
3.8	Discussion	79
4	Multi-fidelity Information Fusion Algorithms for High-Dimensional Systems and Massive Data-sets	82
4.1	Motivation	83
4.2	Machine learning in the frequency domain	87
4.2.1	Bottlenecks of Maximum Likelihood Estimation	87
4.2.2	Frequency-domain sample variogram fitting	88
4.3	Kernel design in high-dimensions	90
4.4	Implementation aspects	95
4.5	Results	97
4.5.1	Borehole function	97
4.5.2	Trid function in 100 dimensions	102
4.5.3	Sobol function in 1,000 dimensions	105
4.6	Discussion	109
5	Modeling Continuum Blood Flow and Biomechanics	112
5.1	Motivation	113
5.2	Nonlinear 1D blood flow models	114
5.2.1	Accounting for arterial viscoelasticity	115
5.2.2	The Discontinuous Galerkin scheme	120
5.2.3	Boundary Conditions	124
5.3	Linear 1D blood flow models	125
5.4	3D blood flow models	127
5.4.1	Flow in rigid vessels	127
5.4.2	Flow-structure interactions	131
5.5	Coupling of the 3D and 1D solvers	133
5.5.1	The coupling algorithm	136
5.5.2	Remarks	138

5.6	Results	142
5.6.1	3D/1D comparison in a straight pipe	143
5.6.2	3D-1D coupling for a straight pipe	146
5.6.3	3D-1D coupling for a Y-shaped bifurcation	149
6	Fractional-order Viscoelasticity in Arterial Biomechanics	155
6.1	Motivation	156
6.2	Viscoelastic Models of the Arterial Wall	159
6.2.1	The Fractional SLS model (Fractional Kelvin-Zener model)	160
6.2.2	Evaluation of Hereditary Integrals	164
6.3	Results	168
6.3.1	Simulation setup	168
6.3.2	Comparison of integer- and fractional-order viscoelastic models	169
6.3.3	Global sensitivity analysis of the stochastic fractional SLS model	174
6.4	Discussion	177
7	A Fractal-tree Closure Model for Simulating Blood Flow in Large Arterial Networks	180
7.1	Motivation	181
7.2	Fractal-tree closure model	184
7.3	Implementation aspects	186
7.4	Case studies and simulation setup	187
7.5	Results	190
7.5.1	Model sensitivity on the cut-off radius	190
7.5.2	Overcoming the cut-off radius sensitivity	192
7.5.3	The Fahraeus-Lindqvist effect	195
7.5.4	Effect of distal outflow boundary conditions	196
7.5.5	3D-1D coupling in a cerebral network with 10 outlets	198
7.6	Discussion	202
8	Calibration of Blood Flow Simulations via Multi-fidelity Surrogates	205
8.1	Motivation	206
8.2	Efficient global optimization	209
8.3	Results	212
8.3.1	Y-shaped bifurcation	212
8.4	Discussion	220
9	Conclusions and Future Work	222
9.1	Conclusions	223
9.2	Future work	224

List of Tables

2.1	<i>Design optimization: Examples of possible design criteria.</i>	11
3.1	<i>Helmholtz equation in 2D: Leading size dimension of the precision Q_u and homogeneous-homogeneous Schur complement $S_{\mathcal{H}\mathcal{H}}$ matrices for different number of expansion modes n.</i>	65
3.2	<i>Stochastic incompressible flow: Size of the uniform discretization grid in the (σ_1, σ_2)-plane, and number of samples per grid-point for each of the probabilistic methods employed.</i>	78
4.1	<i>Sobol function: Breakdown of the total computational cost with respect to the implementation steps outlined in Sec. 4.4.</i>	108
5.1	<i>3D/1D comparison in a straight pipe: Non-dimensional time-step Δt and computation time per cycle for each simulation considered.</i>	143
5.2	<i>3D/1D comparison in a straight pipe: Relative $\mathbb{L}_2$ error at the outlet.</i>	144
5.3	<i>3D-1D coupling: 3D-1D coupling for a straight pipe: Relative $\mathbb{L}_2$ error at the outlet.</i>	147
5.4	<i>3D-1D coupling: 3D-1D coupling for a Y-shaped bifurcation: Relative $\mathbb{L}_2$ error at outlet #1.</i>	152
5.5	<i>3D-1D coupling: 3D-1D coupling for a Y-shaped bifurcation: Relative $\mathbb{L}_2$ error at outlet #2.</i>	152
6.1	<i>Comparison of integer- and fractional-order viscoelastic models: Viscoelastic model parameters.</i>	171
7.1	<i>Fractal tree closure models: Total number of vessels and average number of generations attached to each outlet of the parent CoW and arm networks as a function of the fractal-tree cut-off radius r_{cut}.</i>	190
7.2	<i>3D-1D coupling in a cerebral network with 10 outlets: Simulation properties.</i>	199
7.3	<i>3D-1D coupling in a cerebral network with 10 outlets: Relative errors of the pressure gradient Δp between the inlet and each of the 10 outlets of the 3D domain. We consider $r_{cut} = 100\mu m$ to be our reference case.</i>	200
7.4	<i>3D-1D coupling in a cerebral network with 10 outlets: Relative errors of the flow-rate measured at each of the 10 outlets of the 3D domain. We consider $r_{cut} = 100\mu m$ to be our reference case.</i>	200

8.1	<i>Blood flow in a symmetric Y-shaped bifurcation:</i> Optimal terminal resistances and corresponding relative $\mathbb{L}_2$ -error in the inlet pressure wave over one cardiac cycle, for each iteration of the EGO algorithm. The comparison is done with respect to the inlet pressure wave p_{opt} corresponding to the optimal configuration of resistances predicted by the reference solution obtained using 10,000 high-fidelity samples. . .	219
-----	---	-----

List of Figures

2.1	<i>Schematic representation of the multi-fidelity design under uncertainty framework:</i> the quantity of interest $\mathbf{Y}(\mathbf{x}; \xi)$ is a <i>random field</i> in the design space, realizations of which could originate from information sources of variable-fidelity (such as computer codes, experiments, expert opinion, etc.). A response surface $S(\mathbf{x})$ encodes the dependence of a derived quantity of interest $f(\mathbf{Y}(\mathbf{x}; \xi))$ on the input design variables $\mathbf{x}$ and uncertain parameters ξ . The ability to efficiently construct response surfaces (e.g. $S(\mathbf{x}) = \mathbb{E}[f(\mathbf{Y}(\mathbf{x}; \xi))]$) allows one to assess the system's performance as a function of the random inputs and identify optimal configurations based on the desired design criteria (see Table 2.1).	11
2.2	<i>Gaussian process regression:</i> <i>Left:</i> Samples of a zero-mean Gaussian prior over functions. <i>Middle:</i> Samples of the calibrated Gaussian posterior in view of noise-less data (<i>dashed</i> lines), and the conditional mean of $Z Y$ (<i>solid</i> line). <i>Right:</i> Samples of the calibrated Gaussian posterior in view of noisy data (<i>dashed</i> lines), and the conditional mean of $Z Y$ (<i>solid</i> line). The grey shadowed regions depict the envelope defined by two standard deviations (figure taken from Rasmussen (2006)).	12
2.3	<i>Multi-fidelity in models and in probability space:</i> m models of variable fidelity in physical space are driven by random input, producing a random response surface $\mathbf{Y}_m(\mathbf{x}; \xi)$. For example, the expectation of a derived quantity of interest $\mathbb{E}[f(\mathbf{Y}_m(\mathbf{x}; \xi))]$ can be estimated by employing p methods of variable fidelity in probability space.	18
2.4	<i>Multi-fidelity in models and in probability space:</i> Family of response surfaces resulting from simultaneously addressing multi-fidelity in models and in probability space. Physical model fidelity is increased along the columns (red arrow) and probability space model fidelity increases along the rows (blue arrow). The yellow arrow represents a possible optimal information fusion path in the combined fidelity space.	20
2.5	<i>Multi-fidelity modeling of ocean waves:</i> Exact response surface and co-kriging predictor constructed using 310 low-fidelity and 80 high-fidelity observations.	24
2.6	<i>Multi-fidelity modeling of ocean waves:</i> <i>Left:</i> Point-wise $\mathbb{L}_\infty$ error of the co-kriging predictor (310 low-fidelity and 80 high-fidelity observations). <i>Right:</i> Point-wise variance of the co-kriging predictor. The reference solution was obtained using 2,500 high-fidelity samples.	25

2.7	<i>Multi-fidelity modeling of ocean waves:</i> Visual comparison of the exact solution (2,500 high-fidelity samples) (<i>right</i>), and the co-kriging predictor (310 low-fidelity and 80 high-fidelity observations) (<i>left</i>). . . .	25
2.8	<i>Data-assimilation of coastal data: Left:</i> Kriging predictor (100 noisy observations). <i>Right:</i> Kriging variance.	26
2.9	<i>Data-assimilation of coastal data: Left:</i> Kriging predictor (100 noisy observations).	27
2.10	<i>Data-assimilation of coastal data: Left:</i> Robustness of the riging predictor with respect to corrupted measurements. (<i>Left to right</i>) Predictions $\hat{y}$ versus corrupted data y for increasing noise amplitudes. . .	27
2.11	<i>Effect of low correlation:</i> Ordinary kriging predictor (<i>red</i> dashed line) through 6 equispaced observations (<i>red</i> circles) of $f_h(x)$ (<i>black</i> solid line). The gray shaded area corresponds to two standard deviations of the Gaussian predictor.	30
2.12	<i>Effect of low correlation: Left:</i> Two-level recursive co-kriging predictor for a low-fidelity model $f_l(x)$ with $c = 0.5$. <i>Right:</i> Two-level recursive co-kriging predictor for a low-fidelity model $f_l(x)$ with $c = 0.4$. In both cases, the training data-set consists of 11 equispaced observations of $f_l(x)$, and 6 equispaced observations of $f_h(x)$, while the gray shaded area depicts two standard deviations of the Gaussian predictor. . . .	31
3.1	<i>The spectral/hp element method:</i> Sketch of the unstructured tessellation of the SPDE domain Ω the polynomial basis employed. The solution domain is decomposed into non-overlapping elements; within each element the solution is approximated by vertex, edge, face and (in 3D) interior modes. The shape functions associated with the vertex, edge and face modes for a fourth-order polynomial expansion ($P=4$) defined on triangular and quadrilateral faces are shown in color. . . .	42
3.2	<i>Helmholtz equation in 2D:</i> Exponential convergence rate of the nodal spectral element discretization for a Helmholtz equation in 2D. The error between the numerical and the exact solution has been computed for different numbers of expansion modes n , using the L^∞ , L^2 , and H^1 norms, respectively.	63
3.3	<i>Helmholtz equation in 2D:</i> Condition number of the precision Q_u and homogeneous-homogeneous Schur complement $S_{\mathcal{H}\mathcal{H}}$ for different number of expansion modes n	65
3.4	<i>Helmholtz equation in 2D:</i> Sparsity patterns of the Helmholtz, precision and homogeneous-homogeneous Schur complement matrices for a second-order polynomial expansion basis ($n = 3$).	65
3.5	<i>Helmholtz equation in 2D:</i> Sparsity patterns of the Helmholtz, precision and homogeneous-homogeneous Schur complement matrices for a third-order polynomial expansion basis ($n = 4$).	66
3.6	<i>Helmholtz equation in 2D:</i> Sparsity patterns of the Helmholtz, precision and homogeneous-homogeneous Schur complement matrices for a sixth-order polynomial expansion basis ($n = 7$).	66
3.7	<i>Helmholtz equation in 2D:</i> Observation points, kriging prediction and pointwise L^∞ error for a single observation point located at $(1,0)$. <i>Top to bottom:</i> Nodal spectral element discretization with 4 quadrilateral elements and 3,5,7, and 9 expansion modes, respectively.	68

3.8	<i>Helmholtz equation in 2D:</i> Observation points, kriging prediction and kriging variance for a nodal spectral element discretization with 4 quadrilateral elements and $n = 7$ expansion modes. <i>Top to bottom:</i> Increasing the number of randomly selected observations from 1,5,10, to 15, respectively.	69
3.9	<i>A prototype problem in risk-averse design:</i> Exact response surface $\mathcal{R}_{0.8}(Y)$ and co-kriging predictor constructed using 120 low-fidelity (SG-L2) and 10 high-fidelity (PCM) observations. (x_1^*, x_2^*) denotes the optimal solution to the optimization problem of Eq. 3.63, while the inset plot shows the point-wise absolute error of the co-kriging predictor.	72
3.10	<i>Multi-fidelity regression of a random function:</i> Exact response surface $S = \mathbb{E}[Y(\mathbf{x}; \xi)]$ and co-kriging predictor constructed using four levels of fidelity: 80 S_{11} points ($f_c(\mathbf{x}; \xi)$, SG-L2), 40 S_{12} points ($f_c(\mathbf{x}; \xi)$, SG-L5), 10 S_{21} points ($f_e(\mathbf{x}; \xi)$, SG-L2), and 5 S_{22} points ($f_e(\mathbf{x}; \xi)$, SG-L5). The inset plot shows the point-wise variance of the co-kriging predictor.	75
3.11	<i>Incompressible stochastic flow past a cylinder:</i> (Left) The computational mesh consists of 412 triangular spectral elements of polynomial order 11. (Right) Standard deviation of the stream-wise velocity component at time $t = 4$. The quantity of interest here is the base pressure coefficient C_{BP} , where p_B denotes the fluid pressure on the cylinder surface at 180° from the stagnation point (base pressure), while U_∞ and p_∞ are the free-stream velocity and pressure, respectively.	77
3.12	<i>Incompressible stochastic flow past a cylinder:</i> Probabilistic collocation (PCM) samples of the velocity inflow profile on a uniform 6×5 (σ_1, σ_2) -grid. At each grid point, a total of 64 samples is used to integrate out the randomness introduced by (ξ_1, ξ_2) in Eq. 3.68. Note that σ_1 controls the skewness of the inflow profile, while σ_2 acts as a noise amplifier.	78
3.13	<i>Incompressible stochastic flow past a cylinder:</i> Response surface of the 0.6-superquantile risk of the base pressure coefficient $\mathcal{R}_{0.6}(C_{BP})$, computed using a 3-level recursive co-kriging scheme, on 100 low-fidelity (MC), 15 intermediate-fidelity (SG-L2), and 5 high-fidelity (PCM) observation points in the (σ_1, σ_2) -plane. The exact surface corresponds to 357 uniformly distributed observations of the highest fidelity model (PCM). The inset plot shows the point-wise variance of the co-kriging predictor.	80
4.1	<i>Kernel design in high-dimensions:</i> Sketch of the undirected graph defined by the Sobol sensitivity indices of 5-dimensional function $y(x_1, x_2, x_3, x_4, x_5)$. Here we have one third-order interaction, four second-order interactions and five first-order terms. The size of the lines and the circles indicate the magnitude of the sensitivities. The corresponding maximal cliques are $\mathcal{C}_1 = \{x_1, x_2, x_3\}$, $\mathcal{C}_2 = \{x_1, x_4\}$, $\mathcal{C}_3 = \{x_5\}$	93
4.2	<i>Borehole function:</i> Sketch of the undirected graph of the RS-HDMR sensitivity indices generated using 1,024 observations of $S_{11}(\mathbf{x})$. The first-order sensitivity indices are $S_i = [0.837, 0.0, 0.0, 0.042, 0.0, 0.041, 0.041, 0.0]$, while the non-zero pair-wise sensitivities are $S_{14} = 0.011$, $S_{16} = 0.011$, and $S_{14} = 0.010$, capturing 99.7% of the total variance in the observations.	101

4.3	<i>Borehole function: Left:</i> Density plot of the of the frequency distribution of the exact solution $\mathbb{E}[y]$ (<i>blue</i> solid line), versus the estimated $\mathbb{E}[\hat{y}]$ (dashed lines) resulting from co-kriging and the HDMR representation. The <i>red</i> dashed line corresponds to the final co-kriging predictor accounting for information fusion along the path $S_{11} \rightarrow S_{12} \rightarrow S_{21} \rightarrow S_{22}$. <i>Right:</i> Scatter plot of the exact solution $\mathbb{E}[y]$ (<i>black</i> dashed line), versus the co-kriging predictor $\mathbb{E}[\hat{y}]$ at each level (circles) at 2,000 randomly sampled test locations. The <i>black</i> circles correspond to the final co-kriging predictor accounting for information fusion along the path $S_{11} \rightarrow S_{12} \rightarrow S_{21} \rightarrow S_{22}$	102
4.4	<i>Trid function:</i> Sketch of the undirected graph of the RS-HDMR sensitivity indices generated using 8,196 observations of $y_1(\mathbf{x})$. The structure is in agreement with the definition of the Trid function in Eq. 4.28, suggesting the presence of only pairwise interactions that lead to 99 maximal cliques, each containing two active dimensions.	103
4.5	<i>Trid function: Left:</i> Density plot of the of the frequency distribution of the exact solution $y = y_3(\mathbf{x}^*)$ (<i>blue</i> solid line), versus the estimated $\hat{y}(\mathbf{x}^*)$ (dashed lines) resulting from co-kriging and the HDMR representation. The <i>red</i> dashed line corresponds to the final co-kriging predictor accounting for information fusion along the path $y_1 \rightarrow y_2 \rightarrow y_3$. <i>Right:</i> Scatter plot of the of the of the exact solution $y = y_3(\mathbf{x}^*)$ (<i>black</i> dashed line), versus the co-kriging predictor $\hat{y}(\mathbf{x}^*)$ at each level (circles) at 2,000 test locations, randomly sampled within unit hyper-spheres centered at observations. The <i>black</i> circles correspond to the final co-kriging predictor accounting for information fusion along the path $y_1 \rightarrow y_2 \rightarrow y_3$	105
4.6	<i>Sobol function: Left:</i> Density plot of the of the frequency distribution of the exact solution $y(\mathbf{x}^*)$ (<i>blue</i> solid line), versus the estimated $\hat{y}(\mathbf{x}^*)$ (dashed lines) resulting from kriging and the HDMR representation. <i>Right:</i> Scatter plot of the of the of the exact solution $y(\mathbf{x}^*)$ (<i>black</i> dashed line), versus the kriging predictor $\hat{y}(\mathbf{x}^*)$ (<i>black</i> circles) at 2,000 test locations, randomly sampled within unit hyper-spheres centered at observations.	108
5.1	<i>Nonlinear 1D blood flow:</i> Flow in a 1D compliant artery (from Sherwin et al. (2003)).	114
5.2	<i>Nonlinear 1D blood flow:</i> viscoelastic models considered. <i>Left:</i> Voigt. <i>Right:</i> Kelvin.	116
5.3	<i>Nonlinear 1D blood flow: Left:</i> Sketch of the RCR outflow model. <i>Right:</i> Sketch of the forward and backward characteristics at the 1D-0D coupling interface (figure taken from Alastruey et al. (2008)). . . .	125
5.4	<i>3D-1D coupling:</i> Computational domain for 3D-1D overlapping coupling for a straight pipe.	133
5.5	<i>3D-1D coupling:</i> Sketch of the 3D-1D weak coupling algorithm	136
5.6	<i>3D-1D coupling:</i> Flow-rate interpolation for intermediate asynchronous 1D solver steps.	137
5.7	<i>3D-1D coupling:</i> Normal flux close to the outflow boundary of a straight pipe (<i>left</i>) and a converging pipe with 5 degrees tapering (<i>right</i>).	139
5.8	<i>3D/1D comparison in a straight pipe:</i> Comparison of computed flow-rates at the inlet and outlet.	145

5.9	<i>3D/1D comparison in a straight pipe:</i> Comparison of computed pressures at the inlet and outlet.	145
5.10	<i>3D-1D coupling:</i> 3D-1D coupling for a straight pipe: Comparison of computed flow-rates at the inlet and outlet (<i>RC</i> -coupling, no overlap).	148
5.11	<i>3D-1D coupling:</i> 3D-1D coupling for a straight pipe: Comparison of computed pressures at the inlet and outlet (<i>RC</i> -coupling, no overlap).	148
5.12	<i>3D-1D coupling:</i> 3D-1D coupling for a Y-shaped bifurcation: Considered model for the Y-shaped bifurcation: full 3D domain (left), 3D-1D non-overlapping domain (center), 3D-1D overlapping domain (right). The orange region depicts the overlap of the 3D and 1D domains.	150
5.13	<i>3D-1D coupling:</i> 3D-1D coupling for a Y-shaped bifurcation: Flow-rate oscillations at the 3D-1D non-overlapping interface #1 (<i>RC</i> -coupling).	151
5.14	<i>3D-1D coupling:</i> Comparison of computed flow-rates at the inlet and outlets (<i>RC</i> -coupling, 3D-1D overlap).	153
5.15	<i>3D-1D coupling:</i> Comparison of computed pressures at the inlet and outlets (<i>RC</i> -coupling, 3D-1D overlap).	154
6.1	<i>Simulation setup:</i> Patient-specific cranial network from Grinberg et al. (2011a) and PC-MRI flowrate measurements at the four inlets (L. ICA, R. ICA, L. VA, R. VA)	169
6.2	<i>Comparison of integer- and fractional-order viscoelastic models:</i> Comparison of four viscoelastic model parameter sets: Variability in the computed flowrate at outlets # 23, 30, 44 and 50.	172
6.3	<i>Comparison of integer- and fractional-order viscoelastic models:</i> Comparison of four viscoelastic model parameter sets: Variability in the computed pressure wave and corresponding pressure-area hysteresis loop at the inlet (L. ICA) and distal outlet # 29.	173
6.4	<i>Comparison of integer- and fractional-order viscoelastic models:</i> Computed pressure and pressure-area hysteresis loop at the inlet (L. ICA) for different values of the fractional order $0 \leq \alpha \leq 1$. The dashed line corresponds to the integer-order model SLS1 (recovered for $\alpha = 1.0$) with a wider relaxation spectrum: $\tau_\sigma = 0.025s$, $\frac{\tau_\epsilon}{\tau_\sigma} = 4$	174
6.5	<i>Global sensitivity analysis of the stochastic fractional SLS model:</i> Computed mean flowrate (in <i>black</i>) and a variation of ± 1 standard deviation (in <i>red</i>) at distal outlets # 44 and 50.	175
6.6	<i>Global sensitivity analysis of the stochastic fractional SLS model:</i> Computed mean pressure (in <i>black</i>) and a variation of ± 1 standard deviation (in <i>red</i>) at the inlet (L. ICA) and distal outlet # 50.	176
6.7	<i>Global sensitivity analysis of the stochastic fractional SLS model:</i> Variance decomposition of the computed pressure waveform at the inlet (L. ICA). S_i and S_{ij} are the global sensitivity indices of each random parameter and their pairwise interactions, respectively.	177
7.1	<i>Fractal-tree closure model:</i> <i>Left:</i> Fractal-trees are attached at each outlet of a patient-specific arterial network, resulting to an arterial domain with millions of compliant vessels in which blood flow is modeled using Eq. 5.1. <i>Right:</i> Flow is driven by high resolution, patient-specific PC-MRI measurements at the inlets (right vertebral artery R. VA, left vertebral artery L. VA, right internal carotid artery R. ICA and left internal carotid artery L. ICA).	186

7.2	<i>Case studies and simulation setup:</i> A detailed arterial network of the arm, courtesy of Watanabe et al. (2013) (276 arteries with 105 terminal outlets, 124 bifurcation points, 18 merging junctions and 85 perforator vessels).	189
7.3	<i>Model sensitivity on the cut-off radius:</i> Sensitivity on the cut-off radius r_{cut} : Computed pressure (<i>blue</i>) and flowrate (<i>red</i>) waves for different values of the cut-off radius r_{cut} . <i>Right:</i> Basilar artery (CoW case). <i>Left:</i> Radial artery (arm case).	191
7.4	<i>Overcoming the cut-off radius sensitivity:</i> <i>Left:</i> Decreasing L/r ansatz for different thresholds r^* . <i>Right:</i> Volume of the generated arterial tree as a function of the cut-off radius for different L/r relations (CoW case). The dashed line corresponds to the estimated target volume of $35mL$.	193
7.5	<i>Overcoming the cut-off radius sensitivity:</i> Waveform convergence with decreasing cut-off radius using the L/r ansatz : Computed pressure (<i>blue</i>) and flowrate (<i>red</i>) waves for different values of the cut-off radius r_{cut} . <i>Left:</i> Basilar artery (CoW case). <i>Right:</i> Radial artery (arm case). Inscribed tables contain the relative error ($\mathbb{L}_2$ norm) of the pressure wave as we decrease the cut-off radius (reference cases: $r_{cut} = 20\mu m$ for CoW and $r_{cut} = 10\mu m$ for the arm).	195
7.6	<i>The Fahraeus-Lindqvist effect:</i> Computed pressure waves for different blood rheology models: radius-depended viscosity suggested by Pries et al. (1992) (non-Newtonian, <i>solid line</i>) versus constant blood viscosity (Newtonian, <i>dashed line</i>) <i>Left:</i> Left internal carotid artery (L. ICA, CoW case). <i>Right:</i> Axillary artery (arm case).	196
7.7	<i>Effect of distal outflow boundary conditions:</i> Effect of distal outflow boundary conditions: Computed pressure (<i>solid lines</i>) and flowrate (<i>dashed lines</i>) waves for different types of distal outflow boundary conditions: Absorbing (<i>red</i>), constant pressure (<i>green</i>), and 3-element RCR windkessels (<i>blue</i>). <i>Left:</i> Anterior communicating artery (AcoA, CoW case). <i>Right:</i> Brachial artery (arm case).	198
7.8	<i>3D-1D coupling in a cerebral network with 10 outlets:</i> Network of 20 cranial arteries with 1 inlet (red) and 10 outlets.	198
7.9	<i>3D-1D coupling in a cerebral network with 10 outlets:</i> Comparison of computed pressure at the outlets (left: 3D with 0D-RC outflow, right: 3D with 1D elastic fractal networks).	201
7.10	<i>3D-1D coupling in a cerebral network with 10 outlets:</i> Comparison of computed flow-rates at the outlets (left: 3D with 0D-RC outflow, right: 3D with 1D fractal networks).	201
8.1	<i>Blood flow in a symmetric Y-shaped bifurcation:</i> Flow is driven by a physiological flow-rate wave at the inlet, while outflow boundary conditions are imposed through 3-element windkessel models.	212
8.2	<i>Blood flow in a symmetric Y-shaped bifurcation:</i> Detailed response surface and identification of the minimum error in the inlet systolic pressure as a function of the total resistance parameters imposed at the outlets. The surface is constructed by probing a nonlinear 1D-FSI model on 10,000 uniformly spaced samples in the space of inputs $\mathcal{X}$.	215

8.3	<i>Blood flow in a symmetric Y-shaped bifurcation: Zeroth iteration of the EGO algorithm. Left: Exact solution versus the co-kriging predictor for the inlet systolic pressure error, trained on 100 low-fidelity and 25 high-fidelity observations. Right: Map of the corresponding expected improvement.</i>	216
8.4	<i>Blood flow in a symmetric Y-shaped bifurcation: Fourth iteration of the EGO algorithm. Left: Exact solution versus the co-kriging predictor for the inlet systolic pressure error, trained on 100 low-fidelity and 25 high-fidelity observations. Right: Map of the corresponding expected improvement.</i>	218
8.5	<i>Blood flow in a symmetric Y-shaped bifurcation: Convergence of the inlet pressure waveform for the first 3 EGO iterations. The exact solution corresponds to the reference results obtained from 10,000 high-fidelity samples, defining the target inlet systolic pressure p_s^*.</i>	219
9.1	<i>Deep networks: (a) Sketch of a deep network of vector-valued random fields $f^{(L)} \in \mathbb{R}^m$. In this particular example we have an observation model $y^{(L)} = f^{(L)}$ for each layer, but we can also have situations in which we have hidden (unobserved) layers. The state space variables (e.g., the design variables) are denoted by $x \in \mathbb{R}^n$. The multi-objective response - random field approximating the QoI by $Y(x)$. The deep network allows us to perform information fusion of different information sources by using scalable machine learning algorithms. (b) Recursive co-kriging method as a particular instance of a deep linear Gaussian network.</i>	226

CHAPTER ONE

Introduction

1.1 Data-driven Scientific Computing

This thesis takes aim at two directions. The first direction involves setting the foundations for a “*new*” type of *data-driven scientific computing*, essentially creating a platform for blending multiple information sources of variable fidelity, e.g., experimental data, high-fidelity numerical simulations, expert opinion, etc., towards creating tractable paths to analyzing the response of complex physical and biological systems. Standing on the verge of traditional scientific computing and contemporary statistical learning techniques, we elaborate on a novel multi-fidelity information fusion framework that allows for the seamless integration of surrogate-based optimization and uncertainty quantification, and enables the development of efficient algorithms for data assimilation, design optimization, model inversion, and beyond.

The second direction is focused on addressing open questions in the modeling of the human circulatory system, especially the interplay between blood flow and biomechanics in the brain. Despite great growth in computing power and algorithmic sophistication, *in-silico* modeling of blood flow and arterial mechanics is still limited to truncated arterial domains, a fact that introduces the need for reduced-order models and parametric representations to account for the neglected mesoscale dynamics. The implicit need of such simplified representations gives rise to a series of open questions in cardiovascular mathematics, three of which will be at the focal point of our attention.

In what follows, we provide a summary of how the content of this thesis has been organized.

1.1.1 Part 1: Scalable multi-fidelity information fusion algorithms

In Chapter 2 we put forth a general framework for multi-fidelity information fusion that simultaneously takes into account multi-fidelity in physical models (e.g., direct numerical simulations versus empirical formulas), as well as multi-fidelity approximations in probability space (e.g., sparse grids vs. tensor product multi-element probabilistic collocation). We will describe the building blocks of our construction, namely Gaussian process regression and recursive inference schemes based on autoregressive stochastic modeling. This exposition aims to lay the methodological foundations for the first part of this thesis, provide motivation, and state the challenges that will be addressed in the subsequent chapters. These challenges are primarily centered around two long-standing pathologies in statistical learning, the “big- N ” problem for training surrogates on large data-sets, and the curse of dimensionality.

In Chapter 3 we equip the proposed multi-fidelity information fusion framework with efficient predictive inference algorithms based on multi-level recursive co-kriging and Gaussian Markov random fields (GMRFs). We consider families of GMRFs that arise as solutions to certain stochastic partial differential equations (SPDEs), and we introduce a nodal spectral/ hp element discretization that leads to high-order accurate approximations with sparse inverse covariance matrices. Leveraging this sparsity, we develop efficient parallel solvers for constructing response surfaces of complex stochastic dynamical systems that significantly outperform the $\mathcal{O}(N^3)$ barrier of maximum likelihood estimation techniques. The effectiveness of the new algorithms is demonstrated in numerical examples involving the reconstruction of gappy fields, a prototype problem in *risk-averse* design, the regression of random functions, as well as uncertainty quantification in fluid mechanics involving random

laminar wakes behind circular cylinders.

In Chapter 4 we introduce a methodology for multi-fidelity information fusion and predictive inference in high dimensional input spaces and in the presence of massive data-sets. The proposed methodology further extends the paradigm introduced in Chapter 2, in the context of seamlessly integrating surrogate-based modeling and uncertainty quantification. Scaling to high dimensions is achieved by data-driven dimensionality reduction techniques based on hierarchical functional decompositions and a graph-theoretic approach for constructing custom auto-correlation models for Gaussian process regression. Multi-fidelity information fusion is facilitated through stochastic auto-regressive schemes and frequency-domain machine learning algorithms that scale linearly with the data. The effectiveness of the new algorithms is demonstrated in several benchmark problems involving the regression of multi-fidelity deterministic and stochastic functions in up to 1,000 input dimensions and 10^5 data-points on a standard desktop computer.

1.1.2 Part 2: Open questions in cardiovascular mathematics

In Chapter 5 we present outline the basic tools employed in this this thesis for studying blood flow and arterial biomechanics *in-silico*. In particular, we provide a detailed presentation of one-dimensional (1D) models of blood flow in compliant arteries that will be extensively used in Chapters 6-8. Combining accuracy in predicting key flow features and low computational cost, such models enable a broad range of systematic studies that would otherwise be intractable if one solely relied on expensive three-dimensional formulations. Moreover, we consider possible interface conditions for the coupling of the geometrically heterogeneous 3D and 1D solvers, and test a novel approach based on overlapping domains. These methods set the ground for the

subsequent findings presented Chapters 6-8. There, we provide systematic studies that aim to elucidate three open questions in cardiovascular mathematics.

The first question involves devising reduced-order models that are robust with respect to model misspecification and sensitivity in the input parameters. This is a key question, as model reduction typically introduces a number of parameters that are very hard to estimate in the clinical setting, and increased sensitivity to them jeopardizes the applicability of such simplified models. In Chapter 6 we address this issue in the context of designing accurate and robust constitutive models for the viscoelastic response of soft tissue under stress.

The second open question is related to outflow boundary conditions imposed at the outlets of truncated arterial domains. Typically, this problem is treated using lumped parameter models, although the flow solution is known to exhibit high sensitivity to such parametrizations. In Chapter 7 we present an effective solution to this closure problem by modeling the neglected downstream vasculature as self-similar fractal trees, in which blood flow dynamics are governed by nonlinear 1D conservation laws.

The third question is related to the inevitable step of model parameter calibration that often results to a prohibitive computational cost. Our treatment here leverages directly the developments presented in the first part of this thesis, as we construct a multi-fidelity, surrogate-based framework for model inversion. The proposed methodology targets the accurate construction of high-dimensional response surfaces, and the efficient pursuit to identify global optima while keeping the number of expensive model evaluations at a minimum. To this end, we utilize the machine learning techniques developed throughout Chapters 2- 5, towards exploiting the resulting predictive inference schemes within a Bayesian optimization setting.

CHAPTER TWO

A General Framework for Multi-fidelity Information Fusion

2.1 Motivation

Progress in perceptibly diverse areas of science such as numerical analysis and scientific computing, design optimization, uncertainty quantification, and statistical learning, have started to carve an emerging trend in engineering design, in which decision making becomes increasingly more data-driven rather than merely relying on empirical formulae and expert opinion. A set of versatile tools, ranging from experiments to stochastic simulations can offer invaluable input towards not only performance optimization, but also risk assessment, cost effectiveness and operational aspects of a design. A smart management of these powerful resources is expected to play a vital role, especially in the case of novel, high-performance unconventional designs where simple empirical extrapolation can provide no guarantee of optimality. However, employing ensembles of high-fidelity experiments or computer simulations for performing data-driven design optimization for realistic cases may easily become a task of prohibitive cost. A viable solution to this problem is offered by considering *surrogate modeling*.

A surrogate model can be simply thought of as an intermediate agent that absorbs information coming from a realization of what we consider to be as the true response of a system (up to measurement error), and is able to perform inexpensive predictions as to what the true response might look like in regions where no a-priori information is available. Modern statistical learning techniques have opened the path for constructing sophisticated surrogates that are able to accurately capture intrinsic features of the response of complex engineering systems (see [Forrester et al. \(2008\)](#), [Rasmussen \(2006\)](#), [Hastie et al. \(2009\)](#)). At the heart of surrogate-based modeling we find classical statistical methods, such as Gaussian process regression, members of which are the so-called *kriging* and *co-kriging* predictors (see [Cressie and Cassie](#)

(1993), [Rasmussen \(2006\)](#)).

Initially introduced in geostatistics by the pioneering work of [Matheron \(1963\)](#) and [Kriging \(1951\)](#), kriging has been a flexible and reliable tool in spatial statistics and has extensively served a widespread spectrum of applications, ranging from classical data assimilation in environmental sciences to modern machine learning techniques in robotics. However, it was the seminal work of [Sacks et al. \(1989\)](#) that introduced kriging in the context of design and analysis of computer experiments, and consequently led to a rapidly growing field of application in which kriging has been used as a predictive tool for constructing response surfaces of engineering systems by exploring the spatial correlation between the output of deterministic computer codes. [Simpson et al. \(2008\)](#) provides a comprehensive summary of the evolution and recent advancements of this metamodeling approach in the context of multidisciplinary design optimization.

Of particular importance to our study is the work of [Kennedy and O'Hagan \(2000\)](#) that introduced the use of first-order auto-regressive stochastic models for predicting the output of a complex computer code when fast approximations are available. We recognize that this approach establishes a coherent mathematical framework for blending heterogeneous variable-fidelity information sources, creating a natural setting for multi-fidelity modeling. [Forrester et al. \(2007\)](#) adopted this methodology to build a two-level co-kriging scheme and successfully applied it for optimizing the drag/dynamic pressure ratio of a transonic aircraft wing.

Although the auto-regressive scheme of [Kennedy and O'Hagan \(2000\)](#) has served well many disciplines and applications, it suffers from the same limitations as any kriging/co-kriging method. The biggest practical limitation stems from the need to repeatedly invert large, dense, ill-conditioned covariance matrices at the machine

learning stage, where the metamodel is trained on a set of known observations. To address this issue, we have aligned our work with some new interesting findings that can significantly enhance the computational efficiency of kriging/co-kriging. First, the treatise of [Le Gratiet and Garnier \(2014\)](#) suggests that co-kriging schemes arising from the model of [Kennedy and O’Hagan \(2000\)](#) with s -levels of variable-fidelity information sources can be effectively decoupled, and equivalently formulated in a recursive fashion as s independent kriging problems. This allows for the construction of predictive co-kriging schemes by solving a sequence of simpler kriging problems that involve covariance matrices of smaller dimensions (and potentially lower condition numbers), which can be learned from the data by performing optimization in lower dimensional spaces (compared to the coupled setting of [Kennedy and O’Hagan \(2000\)](#)).

The aim of this chapter is to establish a new mathematical framework for surrogate-based building of response surfaces that simultaneously takes into account multi-fidelity in models (e.g., high-fidelity direct numerical simulations vs. low-fidelity empirical formulas) as well as multi-fidelity in probability space (e.g., high-fidelity tensor product multi-element probabilistic collocation vs low-fidelity sparse grid quadratures). We incorporate elements of statistical learning in an auto-regressive co-kriging methodology to cross-correlate ensembles of multi-fidelity surrogates from which we can accurately and efficiently extract the response of complex non-linear stochastic dynamical systems. This framework targets a seamless integration of surrogate-based optimization and uncertainty quantification, providing a launchpad for contemporary engineering design under uncertainty, inverse problems, and beyond.

This chapter is structured as follows. In [Sec. 2.3](#) we outline the basic mathematical concepts upon which the proposed framework is built. In particular, [Sec.](#)

2.2, 2.3 describe the setting of multi-fidelity modeling and introduces the core concept of Gaussian process regression in the form of kriging, co-kriging and multi-level recursive co-kriging. We then define a general auto-regressive setting under which multi-fidelity in models and multi-fidelity in probability space can be addressed in unison. These methods are the cornerstone of this thesis and provide the basis for the algorithms and results presented in Chapters 3, 4.

2.2 Multi-fidelity modeling

Consider a general system whose response can be evaluated using different models. The response typically depends on a set of design parameters $\mathbf{x} \in \mathbb{R}^n$ and is subject to uncertainties encoded by a set of random parameters, $\xi \in \mathbb{R}^m$. Given a set of design criteria, our goal is to construct the corresponding response surface that quantifies the dependence of a derived quantity of interest, $f(\mathbf{Y}(\mathbf{x}; \xi))$, on the input design variables $\mathbf{x}$ and uncertainties ξ (see Fig. 2.1). Design criteria come in different flavors and generally reflect our priorities and objectives in identifying configurations of the design variables $\mathbf{x}$ that qualify as *optimal*. For example, one may pursue to maximize the performance of a system for a wide range of operating conditions (see Phadke (1995)), identify a design with a probability density function (PDF) that matches a target performance (see Seshadri et al. (2014)), or minimize the *risk* associated with undesirable realizations (see Rockafellar and Royset (2014), and Table 2.1). Although many realistic cases seek designs that satisfy multiple objectives, in what follows we will focus on single-objective optimization problems just to enhance the clarity of our presentation.

The process usually referred to as multi-fidelity modeling, elaborates on efficiently

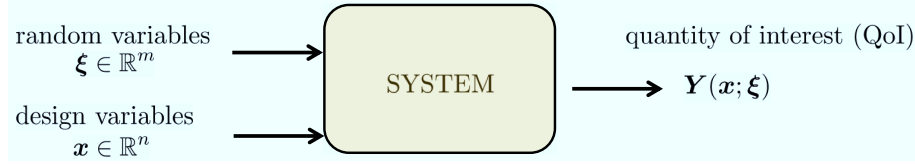


Figure 2.1: *Schematic representation of the multi-fidelity design under uncertainty framework*: the quantity of interest $\mathbf{Y}(\mathbf{x}; \xi)$ is a *random field* in the design space, realizations of which could originate from information sources of variable-fidelity (such as computer codes, experiments, expert opinion, etc.). A response surface $S(\mathbf{x})$ encodes the dependence of a derived quantity of interest $f(\mathbf{Y}(\mathbf{x}; \xi))$ on the input design variables $\mathbf{x}$ and uncertain parameters ξ . The ability to efficiently construct response surfaces (e.g. $S(\mathbf{x}) = \mathbb{E}[f(\mathbf{Y}(\mathbf{x}; \xi))]$) allows one to assess the system's performance as a function of the random inputs and identify optimal configurations based on the desired design criteria (see Table 2.1).

Design criteria	
$\max_{\mathbf{x}} \mathbb{E}[f(\mathbf{Y}(\mathbf{x}; \xi))]$ and $\min_{\mathbf{x}} \text{Var}[f(\mathbf{Y}(\mathbf{x}; \xi))]$	(Phadke (1995))
$\min_{\mathbf{x}} \ p(\mathbf{Y}; \mathbf{x}) - q(\mathbf{Y})\ $, where $q(\mathbf{Y})$ is the PDF of a <i>target performance</i>	(Seshadri et al. (2014))
$\min_{\mathbf{x}} \mathcal{R}[f(\mathbf{Y}(\mathbf{x}; \xi))]$, where $\mathcal{R}$ is a <i>risk measure</i>	(Rockafellar and Royset (2014))

Table 2.1: *Design optimization*: Examples of possible design criteria.

constructing response surfaces by correlating surrogate models of different fidelity. On one side of the fidelity spectrum one may have cheaper surrogate models, which are fast to compute but less trustworthy (e.g. potential flow solvers, empirical formulas, etc), while on the other side we have high-fidelity models that enjoy our outmost trust but can be very expensive to compute (e.g. direct numerical simulations, experiments, etc). In many cases, exploring the correlation between the two allows us to efficiently construct an accurate representation of the response surface by performing relatively few evaluations of an expensive high-fidelity model and more evaluations of a cheaper surrogate. A linking mechanism that undertakes the task of information fusion naturally arises in the context of multivariate Gaussian regression. Next, we introduce this concept in the context of ordinary kriging and co-kriging.

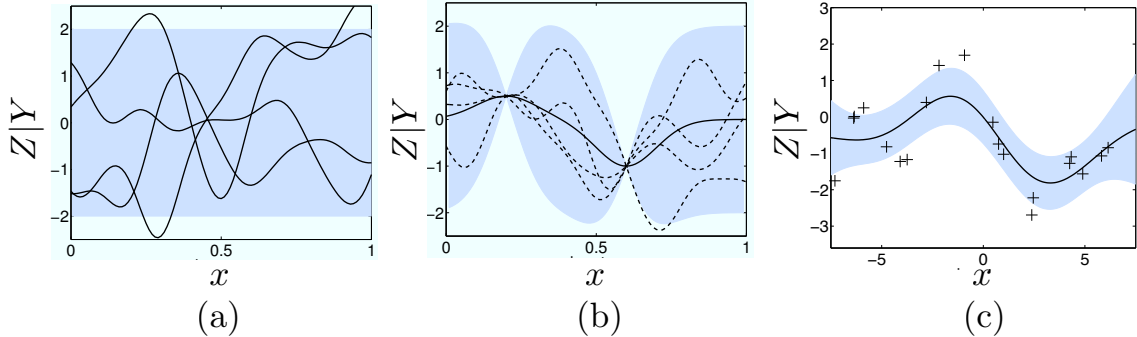


Figure 2.2: *Gaussian process regression*: *Left*: Samples of a zero-mean Gaussian prior over functions. *Middle*: Samples of the calibrated Gaussian posterior in view of noise-less data (*dashed* lines), and the conditional mean of $Z|Y$ (*solid* line). *Right*: Samples of the calibrated Gaussian posterior in view of noisy data (*dashed* lines), and the conditional mean of $Z|Y$ (*solid* line). The grey shadowed regions depict the envelope defined by two standard deviations (figure taken from [Rasmussen \(2006\)](#)).

2.3 Methods

The basic building block of the proposed multi-fidelity information fusion framework is Gaussian process regression (GPR). One way of viewing GPR is as defining a prior distribution over functions, which is then calibrated in view of data using an appropriate likelihood function, resulting to a posterior distribution with predictive capabilities (see Fig. 2.2). In what follows we provide an overview of the key steps in this construction, in the context of the ordinary kriging and recursive co-kriging predictors. The reader is referred to [Rasmussen \(2006\)](#) for a detailed exposition to the subject.

2.3.1 Kriging

The main idea here is to model N scattered observations y of a quantity of interest (QoI) $Y(\mathbf{x})$ as a realization of a Gaussian random field $Z(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$. The observations could be deterministic or stochastic in nature and may well be corrupted by

modeling errors or measurement noise $\mathcal{E}(\mathbf{x})$, which is thereby assumed to be a zero-mean Gaussian random field, i.e $\mathcal{E}(\mathbf{x}) \sim \mathcal{N}(0, \sigma_\epsilon^2 I)$. Therefore, we have the following observation model

$$Y(\mathbf{x}) = Z(\mathbf{x}) + \mathcal{E}(\mathbf{x}). \quad (2.1)$$

The prior distribution on $Z(\mathbf{x})$ is completely characterized by a mean $\mu(\mathbf{x}) = \mathbb{E}[Z(\mathbf{x})]$ and covariance $\kappa(\mathbf{x}, \mathbf{x}'; \theta)$ function, where θ is a vector of hyper-parameters. Typically, the choice of the prior reflects our belief on the structure, regularity, and other intrinsic properties of the quantity of interest $Y(\mathbf{x})$. However, our primary goal here is not just drawing random fields from the prior but to incorporate the knowledge contained in the observations y in order to reconstruct the field $Y(\mathbf{x})$. This can be achieved by computing the conditional distribution $\pi(\hat{y}|y, \theta)$, where $\hat{y}(\mathbf{x}^*)$ contains the predicted values for $Y(\mathbf{x})$ at a new set of locations $\mathbf{x}^*$. If a Gaussian prior is assumed on the hyper-parameters θ then $\pi(\hat{y}|y, \theta)$ is obviously Gaussian and provides a predictive scheme for the estimated values $\hat{y}$. Once $Z(\mathbf{x})$ has been trained on the observed data (see Sec. ??), its calibrated mean $\hat{\mu}$, variance $\hat{\sigma}^2$, and noise variance $\hat{\sigma}_\epsilon^2$ are known and can be used to evaluate the predictions $\hat{y}$, as well as to quantify the prediction variance s^2 as (see [Jones \(2001\)](#) for a derivation),

$$\hat{y}(\mathbf{x}^*) = \hat{\mu} + r^T (R + \hat{\sigma}_\epsilon^2 I)^{-1} (y - \mathbf{1}\hat{\mu}), \quad (2.2)$$

$$s^2(\mathbf{x}^*) = \hat{\sigma}^2 \left[1 - r^T (R + \hat{\sigma}_\epsilon^2 I)^{-1} r + \frac{[1 - r^T (R + \hat{\sigma}_\epsilon^2 I)^{-1} r]^2}{\mathbf{1}^T (R + \hat{\sigma}_\epsilon^2 I)^{-1} \mathbf{1}} \right], \quad (2.3)$$

where $R = \kappa(\mathbf{x}, \mathbf{x}'; \theta)$ is the $N \times N$ correlation matrix of $Z(\mathbf{x})$, $r = \kappa(\mathbf{x}, \mathbf{x}^*; \theta)$ is a $1 \times N$ vector containing the correlation between the prediction and the N training

points, and $\mathbf{1}$ is a $1 \times N$ vector of ones. This is a linear regression scheme known as the best linear unbiased predictor (BLUP) in the statistics literature (see [Cressie and Cassie \(1993\)](#)). Note, that for $\sigma_\epsilon^2 = 0$ the predictor exactly interpolates the training data y , returning zero variance at these locations.

2.3.2 Recursive co-kriging

We provide a brief overview of the recursive co-kriging scheme recently put forth by [Le Gratiet and Garnier \(2014\)](#) - a more efficient version of the well-known auto-regressive inference scheme proposed by [Kennedy and O'Hagan \(2000\)](#) in the context of predicting the output from a complex computer code when fast approximations are available. To this end, suppose that we have s levels of information sources producing outputs $y_t(\mathbf{x}_t)$, at locations $\mathbf{x}_t \in D_t \subseteq \mathbb{R}^d$, sorted by increasing order of fidelity and modeled by Gaussian processes $Z_t(\mathbf{x})$, $t = 1, \dots, s$. Then, the auto-regressive scheme of [Kennedy and O'Hagan \(2000\)](#) reads as

$$Z_t(\mathbf{x}) = \rho_{t-1}(\mathbf{x})Z_{t-1}(\mathbf{x}) + \delta_t(\mathbf{x}), \quad t = 2, \dots, s, \quad (2.4)$$

where $\delta_t(\mathbf{x})$ is a Gaussian field independent of $\{Z_{t-1}, \dots, Z_1\}$ and distributed as $\delta_t \sim \mathcal{N}(\mu_{\delta_t}, \sigma_t^2 R_t)$. Also, $\rho(\mathbf{x})$ is a scaling factor that quantifies the correlation between $\{Z_t(\mathbf{x}), Z_{t-1}(\mathbf{x})\}$. In the Bayesian setting, $\rho(\mathbf{x})$ is treated as a random field with an assigned prior distribution that is later calibrated to the data. Here, to simplify the presentation we assume that ρ is a deterministic scalar, independent of $\mathbf{x}$, and learned from the data through maximum likelihood estimation (see Sec. ??).

The key idea put forth by [Le Gratiet and Garnier \(2014\)](#) is to replace the Gaus-

sian field $Z_{t-1}(\mathbf{x})$ in Eq. 3.55 with a Gaussian field $\tilde{Z}_{t-1}(\mathbf{x})$ that is conditioned on all known observations $\{y_{t-1}, y_{t-2}, \dots, y_1\}$ up to level $(t-1)$, while assuming that the corresponding experimental design sets D_i , $i = 1, \dots, t-1$ have a nested structure, i.e. $D_1 \subseteq D_2 \subseteq \dots \subseteq D_{t-1}$. This essentially allows to decouple the s -level auto-regressive co-kriging problem to s independent kriging problems that can be efficiently computed and are guaranteed to return a predictive mean and variance that is identical to the coupled scheme of Kennedy and O'Hagan (2000). To underline the advantages of this approach, note that the scheme of Kennedy and O'Hagan requires inversion of covariance matrices of size $\sum_{t=1}^s N_t \times \sum_{t=1}^s N_t$, where N_t is the number of observed training points at level t . In contrast, the recursive co-kriging approach involves the inversion of s covariance matrices of size $N_t \times N_t$, $t = 1, \dots, s$.

Once $Z_t(\mathbf{x})$ has been trained on the observed data $\{y_t, y_{t-1}, \dots, y_1\}$ (see Sec. ??), the optimal set of hyper-parameters $\{\hat{\mu}_t, \hat{\sigma}_t^2, \hat{\sigma}_{\epsilon_t}^2, \hat{\rho}_{t-1}, \hat{\theta}_t\}$ is known and can be used to evaluate the predictions $\hat{y}_t$, as well as to quantify the prediction variance s_t^2 at all points in $\mathbf{x}_t^*$ (see Le Gratiet and Garnier (2014) for a derivation),

$$\hat{y}_t(\mathbf{x}_t^*) = \hat{\mu}_t + \hat{\rho}_{t-1} \hat{y}_{t-1}(\mathbf{x}_t^*) + r_t^T (R_t + \hat{\sigma}_{\epsilon_t}^2 I)^{-1} [y_t(\mathbf{x}_t) - \mathbf{1} \hat{\mu}_t - \hat{\rho}_{t-1} \hat{y}_{t-1}(\mathbf{x}_t)], \quad (2.5)$$

$$s_t^2(\mathbf{x}_t^*) = \hat{\rho}_{t-1}^2 s_{t-1}^2(\mathbf{x}_t^*) + \hat{\sigma}_t^2 \left[1 - r_t^T (R_t + \hat{\sigma}_{\epsilon_t}^2 I)^{-1} r_t + \frac{[1 - r_t^T (R_t + \hat{\sigma}_{\epsilon_t}^2 I)^{-1} r_t]^2}{\mathbf{1}_t^T (R_t + \hat{\sigma}_{\epsilon_t}^2 I)^{-1} \mathbf{1}_t} \right], \quad (2.6)$$

where $R_t = \kappa_t(\mathbf{x}_t, \mathbf{x}'_t; \hat{\theta}_t)$ is the $N_t \times N_t$ correlation matrix of $Z_t(\mathbf{x})$, $r_t = \kappa_t(\mathbf{x}_t, \mathbf{x}_t^*; \hat{\theta}_t)$ is a $1 \times N_t$ vector containing the correlation between the prediction and the N_t training points, and $\mathbf{1}_t$ is a $1 \times N_t$ vector of ones. Note that for $t = 1$ the above scheme reduces to the ordinary kriging predictor of Eq. 3.53, 2.3. Also, $\kappa_t(\mathbf{x}_t, \mathbf{x}'_t; \theta_t)$ is the auto-correlation kernel that quantifies spatial correlations at level t .

We recognize that such recursive auto-regressive schemes can provide a rigorous and tractable workflow for multi-fidelity information fusion. This suggests a general framework that targets the seamless integration of surrogate-based prediction/optimization and uncertainty quantification, allowing one to simultaneously address multi-fidelity in physical models (e.g., direct numerical simulations vs. experiments) as well as multi-fidelity in probability space (e.g., the sparse grid quadratures of [Novak and Ritter \(1996\)](#) vs. the multi-element probabilistic collocation of [Foo et al. \(2008\)](#)).

2.3.3 Maximum likelihood estimation

Estimating the hyper-parameters requires learning the optimal set of $\{\mu_t, \sigma_t^2, \sigma_{\epsilon_t}^2, \rho_{t-1}, \theta_t\}$ from all known observations $\{y_t, y_{t-1}, \dots, y_1\}$ at each inference level t . In what follows we will confine the presentation to maximum likelihood estimation (MLE) procedures for the sake of clarity. However, in the general Bayesian setting all hyper-parameters are assigned with prior distributions, and inference is performed via more costly marginalization techniques, typically using Markov Chain Monte Carlo integration (see [Rasmussen \(2006\)](#), [Hastie et al. \(2009\)](#)).

Parameter estimation via MLE at each inference level t is achieved by minimizing the negative log-likelihood of the observed data y_t ,

$$\begin{aligned} \min_{\{\mu_t, \sigma_t^2, \sigma_{\epsilon_t}^2, \rho_{t-1}, \theta_t\}} \quad & \frac{n}{2} \log(\sigma_t^2) + \frac{1}{2} \log |R_t(\theta_t) + \sigma_{\epsilon_t}^2 I| + \\ & + \frac{1}{2\sigma_t^2} [y_t(\mathbf{x}_t) - \mathbf{1}_t \mu_t - \rho_{t-1} \hat{y}_{t-1}(\mathbf{x}_t)]^T [R_t(\theta_t) + \sigma_{\epsilon_t}^2 I]^{-1} [y_t(\mathbf{x}_t) - \mathbf{1}_t \mu_t - \rho_{t-1} \hat{y}_{t-1}(\mathbf{x}_t)], \end{aligned} \quad (2.7)$$

where we have highlighted the dependence of the correlation matrix R_t on the hyperparameters θ_t . Setting the derivatives of this expression to zero with respect to μ_t, ρ_{t-1} and σ_t^2 , we can express the optimal values of $\hat{\mu}_t, \hat{\rho}_{t-1}$ and $\hat{\sigma}_t^2$ as functions of the correlation matrix $(R_t + \sigma_{\epsilon_t}^2 I)$,

$$(\hat{\mu}_t, \hat{\rho}_{t-1}) = [\mathbf{h}_t^T (R_t + \sigma_{\epsilon_t}^2 I)^{-1} \mathbf{h}_t]^{-1} \mathbf{h}_t^T (R_t + \sigma_{\epsilon_t}^2 I)^{-1} y_t(\mathbf{x}_t) \quad (2.8)$$

$$\hat{\sigma}_t^2 = \frac{1}{c} \left\{ [y_t(\mathbf{x}_t) - \mathbf{1}_t \hat{\mu}_t - \hat{\rho}_{t-1} \hat{y}_{t-1}(\mathbf{x}_t)]^T [R_t + \sigma_{\epsilon_t}^2 I]^{-1} [y_t(\mathbf{x}_t) - \mathbf{1}_t \hat{\mu}_t - \hat{\rho}_{t-1} \hat{y}_{t-1}(\mathbf{x}_t)] \right\}, \quad (2.9)$$

where $\mathbf{h}_t = [\mathbf{1}_t \quad \hat{y}_{t-1}(\mathbf{x}_t)]$, and $c = \begin{cases} N_t - 1, & t = 1 \\ N_t - 2, & t > 1 \end{cases}$. Finally, the optimal $\{\hat{\sigma}_{\epsilon_t}^2, \hat{\theta}_t\}$ can be estimated by minimizing the concentrated restricted log-likelihood

$$\min_{\{\sigma_{\epsilon_t}^2, \theta_t\}} \log |R_t(\theta_t) + \sigma_{\epsilon_t}^2 I| + c \log(\hat{\sigma}_t^2). \quad (2.10)$$

2.4 Multi-fidelity in models and in probability space

We can build further upon the presented co-kriging framework to formulate a general methodology that can simultaneously address multi-fidelity in physical models as well as multi-fidelity in probability space. As it is often the case in realistic design scenarios, the output of a system may well be sensitive to a set of inputs ξ that exhibit random variability. Consequently, decision making towards identifying an optimal design is typically informed by exploring the measures of uncertainty that describe the response of the underlying stochastic dynamical system. This response

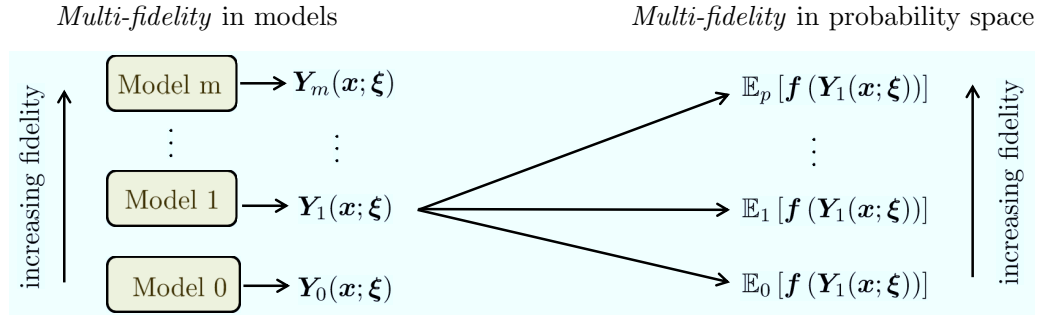


Figure 2.3: *Multi-fidelity in models and in probability space*: m models of variable fidelity in physical space are driven by random input, producing a random response surface $\mathbf{Y}_m(\mathbf{x}; \xi)$. For example, the expectation of a derived quantify of interest $\mathbb{E}[f(\mathbf{Y}_m(\mathbf{x}; \xi))]$ can be estimated by employing p methods of variable fidelity in probability space.

is often characterized by non-Gaussian statistics that can be estimated numerically by utilizing appropriate sampling and integration techniques. The potential non-Gaussianity in the system response should not be confused with the Gaussian nature of the kriging/co-kriging predictors. The former is an inherent property of the dynamical system that generates the observed data, while the later introduces a modeling framework for information fusion.

Similarly to having multi-fidelity in models, methods of different fidelity can also be incorporated in probability space to provide an accurate quantification of uncertainty introduced by random input. This structure is schematically illustrated in Fig. 2.3, where m models of variable fidelity in physical space are driven by random input, hence producing a random response surface $\mathbf{Y}_m(\mathbf{x}; \xi)$. In return, any uncertainty quantification measure of $\mathbf{Y}_m(\mathbf{x}; \xi)$, such as for e.g. the expectation $\mathbb{E}[\mathbf{Y}_m(\mathbf{x}; \xi)]$ or the risk $\mathcal{R}[\mathbf{Y}_m(\mathbf{x}; \xi)]$, can be estimated using a set of p variable-fidelity methods, such as Monte Carlo integration (see [Binder and Heermann \(2010\)](#)) or probabilistic multi-element collocation on tensor product grids (see [Foo et al. \(2008\)](#), and Fig. 2.3).

This construction results in a family of response surfaces that can be organized hierarchically in a $p \times m$ matrix, where physical model fidelity is increased along the columns and probability space model fidelity increases along the rows (see Fig. 2.4). Then, it is meaningful to allow information fusion along the $\{\rightarrow\}$, $\{\downarrow\}$, $\{\uparrow\rightarrow\}$, $\{\downarrow\rightarrow\}$ directions by employing the autoregressive co-kriging framework presented in Sec. 2.2, 2.3.2. For example, moving along the purely vertical direction $\{\downarrow\}$ results to the following autoregressive expression for the expectation $\mathbb{E}[f(\mathbf{Y}_m(\mathbf{x}; \xi))]$:

$$\mathbb{E}_{k+1}[f(\mathbf{Y}_l(\mathbf{x}; \xi))] = \rho_{k+1}\mathbb{E}_k[f(\mathbf{Y}_l(\mathbf{x}; \xi))] + \delta_{k+1}(\mathbf{x}), \quad k \leq p, \quad l \leq m, \quad (2.11)$$

where the k -index increases with the fidelity of the estimator of $\mathbb{E}[f(\mathbf{Y}(\mathbf{x}; \xi))]$ in probability space, while the l -index increases with model fidelity in physical space.

This structure gives rise to the very interesting task of identifying an optimal path traversal between different models for building an accurate representation of the target response surface. This is an open question that we plan to address in a future study. A possible way of attacking this problem is through stochastic dynamic programming techniques for guiding an optimal allocation of available computational resources (see Royset (2013)). Alternatively, one could employ random graph theory to identify optimal information/entropy diffusion paths, where each graph node is weighted by the fidelity and corresponding cost of each model, while edge weights represent the degree of correlation between different models.

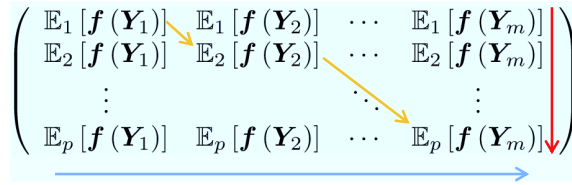


Figure 2.4: *Multi-fidelity in models and in probability space*: Family of response surfaces resulting from simultaneously addressing multi-fidelity in models and in probability space. Physical model fidelity is increased along the columns (red arrow) and probability space model fidelity increases along the rows (blue arrow). The yellow arrow represents a possible optimal information fusion path in the combined fidelity space.

2.5 Challenges to be addressed

2.5.1 Bottlenecks of learning from large data-sets

The computational cost of calibrating model hyper-parameters through MLE is dominated by the inversion of correlation matrices $(R_t + \sigma_{\epsilon_t}^2 I)^{-1}$ at each iteration of the minimization procedure in Eq. 2.10. The inversion is typically performed using the Cholesky decomposition that scales as $\mathcal{O}(N_t^3)$, leading to a severe bottleneck in the presence of moderately big data-sets. This is typically the case for high-dimensional problems where abundance of data is often required for performing meaningful inference. This pathology is further amplified in cases where the noise variance $\sigma_{\epsilon_t}^2$ is negligible and/or the observed data points are tightly clustered in space. Such cases introduce ill-conditioning that may well jeopardize the feasibility of the inversion as well as pollute the numerical solution with errors. Moreover, if an anisotropic correlation kernel $\kappa_t(\mathbf{x}_t, \mathbf{x}'_t; \theta_t)$ is assumed, then the vector of correlation lengths θ_t is d -dimensional, leading to an increasingly complex optimization problem (see Eq. 2.10) as the dimensionality of the input variables $\mathbf{x}_t$ increases. These shortcomings render the learning process intractable for large data-sets, and suggest seeking alternative routes to parameter estimation.

In Chapter 3 we introduce a new inference methodology that exploits the solution of certain stochastic partial differential equations, introducing Gaussian Markov random fields with sparse inverse covariances. Essentially, this enables the design of algorithms that leverage this sparsity, leading to a significant speed up of the inference process.

2.5.2 Bottlenecks of scaling to high dimensions

In general, the design of predictive inference schemes in high-dimensions suffers from the well known curse of dimensionality, as the number of points needed to explore the input space in its entirety increases exponentially with the dimension. This implicit need for big data introduces a severe deadlock for scalability in machine learning algorithms as they often involve the repeated inversion of covariance matrices that quantify the spatial cross-correlations in the observations. Again, this leads to the so called “big N ” problem - an expression used to characterize the demanding operational count associated with handling data-sets comprised of N observations ($N > 1000$). The implications of such large data-sets on learning algorithms are well known, leading to a $\mathcal{O}(N^3)$ scaling for implementations based on maximum likelihood estimation.

Another problem arising in high-dimensions stems from the the common strategy for constructing auto-correlation models by taking the product of one-dimensional autocorrelation kernels. This typically results to an anisotropic covariance model, which assumes that all dimensions actively interact with each other. However, as the dimensionality is increased, one would hope to find sparsity in the input space, i.e. dimensions with negligible or very weak pairwise interactions.

In Chapter 4 we will present a set of new algorithms that introduce a tractable solution path in high dimensional input spaces and in the presence of massive data-sets.

2.6 Demonstration

Here we present two demonstrations to highlight the applicability of the proposed framework to problems of different nature. The first benchmark involves the accurate reconstruction of the response predicted by variable fidelity partial differential equation models that describe the dynamics of deep water waves. The analysis is based on the recent work of [Cousins and Sapsis \(2015a,b\)](#), [Mohamad et al. \(2015\)](#). The second example demonstrates the use of the presented methods for data-assimilation of noisy field measurements for the reconstruction of the bathymetry map in a coastal region. Finally, we present a representative case in one input dimension, where the proposed multi-fidelity formulation fails to result to an improved predictor due to lack of adequate correlations between different fidelity levels.

2.6.1 Multi-fidelity modeling of ocean waves

Understanding and predicting large ocean waves is critical due to the catastrophic damage they inflict on ships and other coastal structures. However, understanding these waves is challenging: due to the large amplitudes of the waves, their dynamics are nonlinear. To better understand these extreme waves, we model the behavior of wave groups. Some wave groups will focus, increasing in amplitude, while others disperse and decrease in amplitude. We wish to quantify how this behavior depends

on the amplitude A and length scale L of the group. Specifically, we determine whether or not a particular group will focus and, if so, the corresponding increase in amplitude.

High fidelity, nonlinear partial differential equation (PDE) models reproduce these extreme wave group dynamics well, but are expensive to solve. Thus it is computationally advantageous to use reduced order models. These reduced order models qualitatively reproduce the true dynamics but have error of up to 30% for some groups. Using the proposed multi-fidelity information fusion framework allows us to approximate the true dynamics by combining a small number of simulations of the high fidelity model with many simulations of the reduced order model.

The quantity of interest here is the spatio-temporal maximum of the wave envelope amplitude u normalized by the initial wave amplitude, i.e., $Y = \max |u|/A_0$. This quantity is 1 when defocusing occurs and the amplitude decreases. Values of this ratio larger than 1 indicate that the associated group focuses, increasing in amplitude. For more details on this modeling approach the reader is referred to the recent work of [Cousins and Sapsis \(2015a,b\)](#), [Mohamad et al. \(2015\)](#).

In [Fig. 2.5](#) we present the response surface of the quantity of interest Y as a function of the input variables $\mathbf{x} = (A_0, L_0)$, where L_0 is the initial wavelength. The response has been reconstructed using 80 samples of an expensive to compute high-fidelity PDE model supplemented with 310 observations coming from a reduced order low-fidelity model. We observe that the resulting multi-fidelity co-kriging predictor is in good agreement with the “exact” solution obtained using 2,500 samples of the high-fidelity code. The accuracy of the prediction is demonstrated in [Fig. 2.6\(a\)](#), where we plot the point-wise error in the $\mathbb{L}_\infty$ norm, while [Fig. 2.6\(b\)](#) shows the spatial distribution of the predictor’s uncertainty. Finally, in [Fig. 2.7](#) we present a

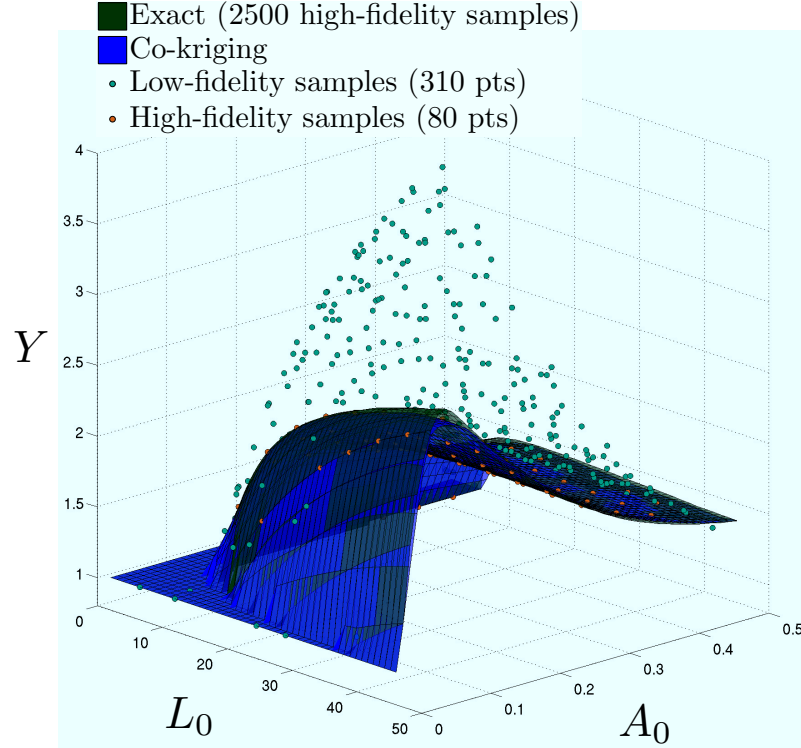


Figure 2.5: *Multi-fidelity modeling of ocean waves*: Exact response surface and co-kriging predictor constructed using 310 low-fidelity and 80 high-fidelity observations.

visual comparison of the exact solution versus the co-kriging surrogate. It is evident that the proposed multi-fidelity analysis can accurately capture the transition in the dynamics that results to a jump in the response surface, but also result to orders of magnitude in speed-up for obtaining an accurate representation.

2.6.2 Data-assimilation of coastal bathymetry data

Our objective here it to use reconstruct the bathymetry field of a coastal region from noisy measurements. To this end, we have available 1,321 depth measurements (courtesy of Mike Sacarny, MIT) and we will test the robustness of the kriging predictive distribution with respect to corrupting the data with additional Gaussian noise. In Fig. 2.8(a),(b) we present the kriging prediction and variance, respectively, using a

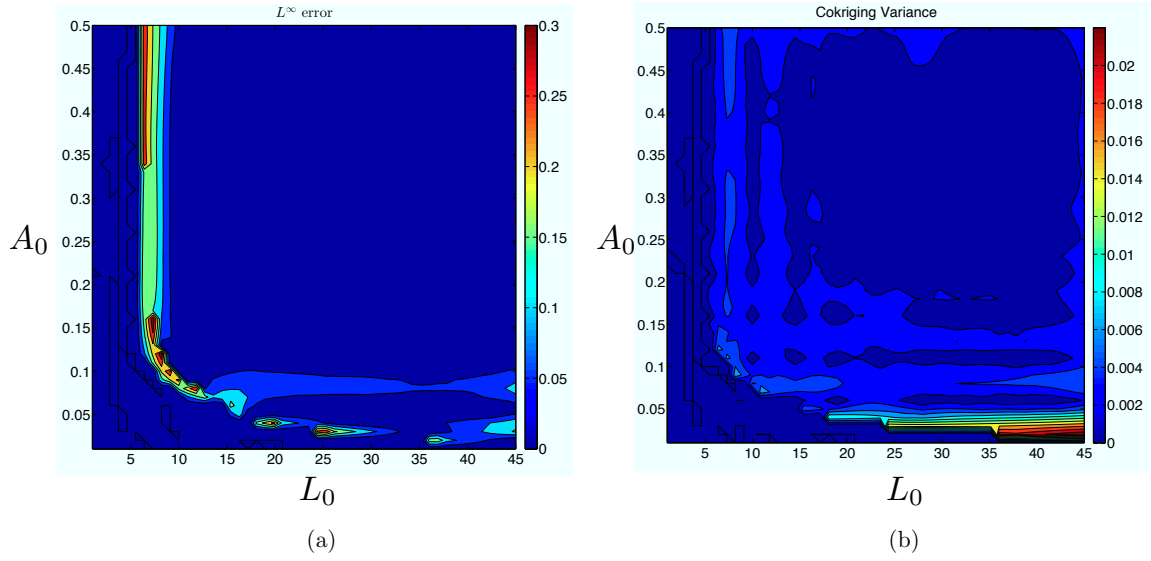


Figure 2.6: *Multi-fidelity modeling of ocean waves*: Left: Point-wise $\mathbb{L}_\infty$ error of the co-kriging predictor (310 low-fidelity and 80 high-fidelity observations). Right: Point-wise variance of the co-kriging predictor. The reference solution was obtained using 2,500 high-fidelity samples.

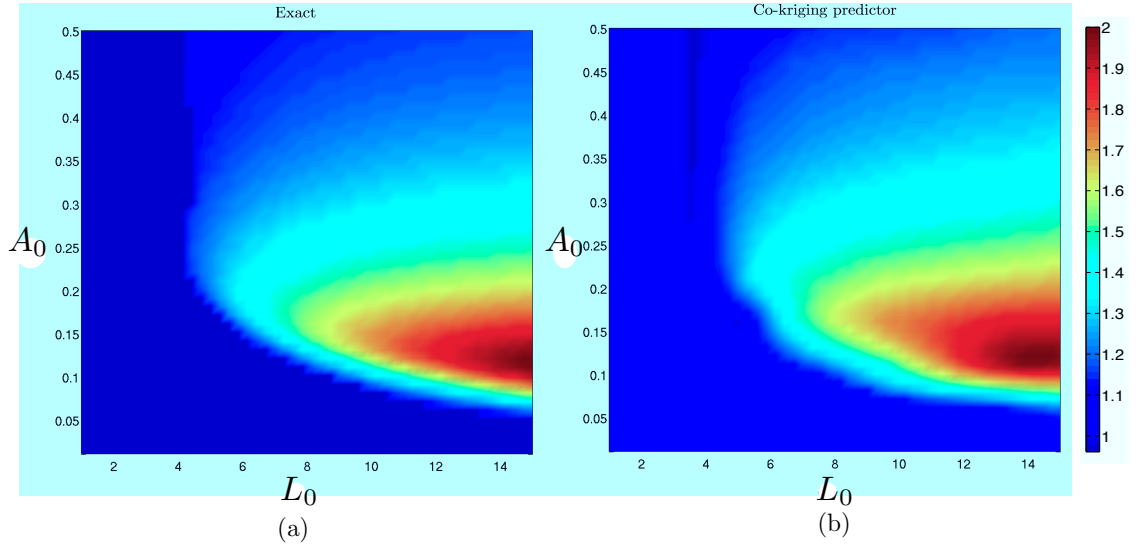


Figure 2.7: *Multi-fidelity modeling of ocean waves*: Visual comparison of the exact solution (2,500 high-fidelity samples) (right), and the co-kriging predictor (310 low-fidelity and 80 high-fidelity observations) (left).

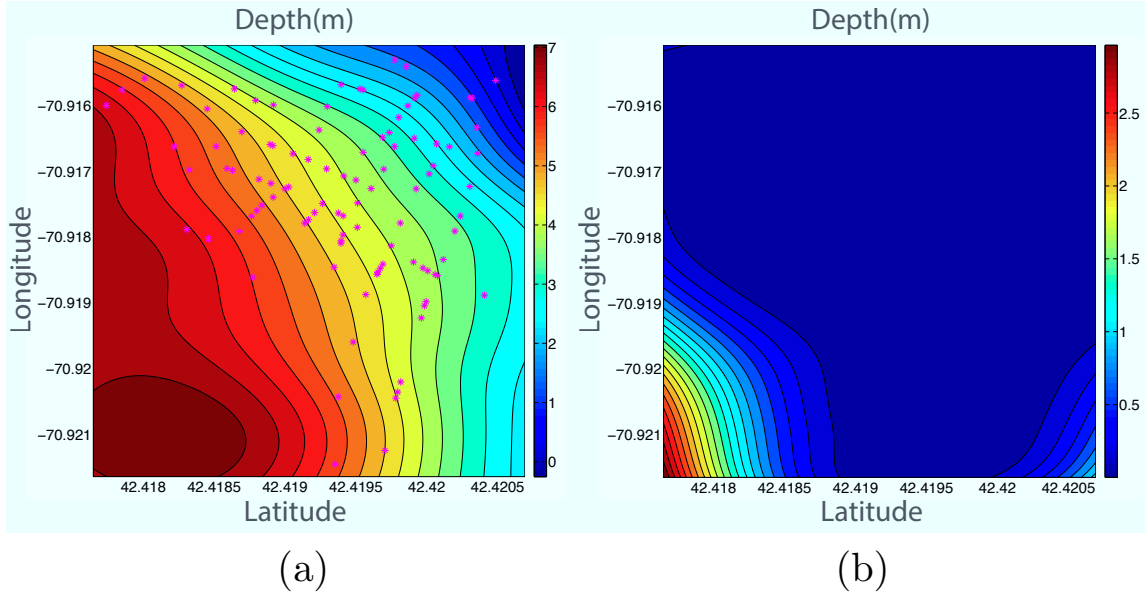


Figure 2.8: *Data-assimilation of coastal data: Left: Kriging predictor (100 noisy observations). Right: Kriging variance.*

5/2 Matérn covariance model, and 100 observations that may contain measurement noise. We observe that this results to a smooth bathymetry map that is in agreement with the exact topology of the region (see Fig. 2.10), while the kriging uncertainty is negligible, except for a region far away from the observations (see bottom right corner of Fig. 2.8(b)).

Finally, we assess the robustness of the kriging scheme with respect to adding a Gaussian perturbation in the original measurements. In Fig. 2.10 we present an illustration of the predictive capacity of the kriging predictor $\hat{y}$ as a function of the noise amplitude. In all cases, the kriging predictor was able to filter out the noise components and return a smooth response surface.

These demonstrations serve as motivation of some of the capabilities that the proposed framework has to offer. In Chapter 3 we provide a detailed overview of the methods employed to produce the aforementioned results, as well as a number of benchmark problems that further illustrate the potential applications of this work.

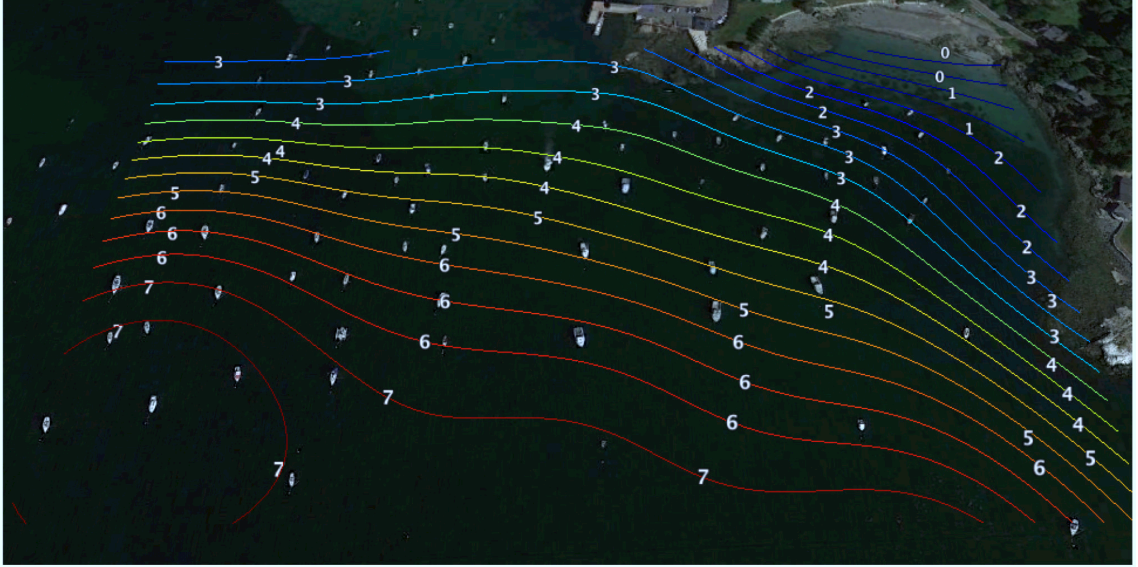


Figure 2.9: *Data-assimilation of coastal data: Left: Kriging predictor (100 noisy observations).*

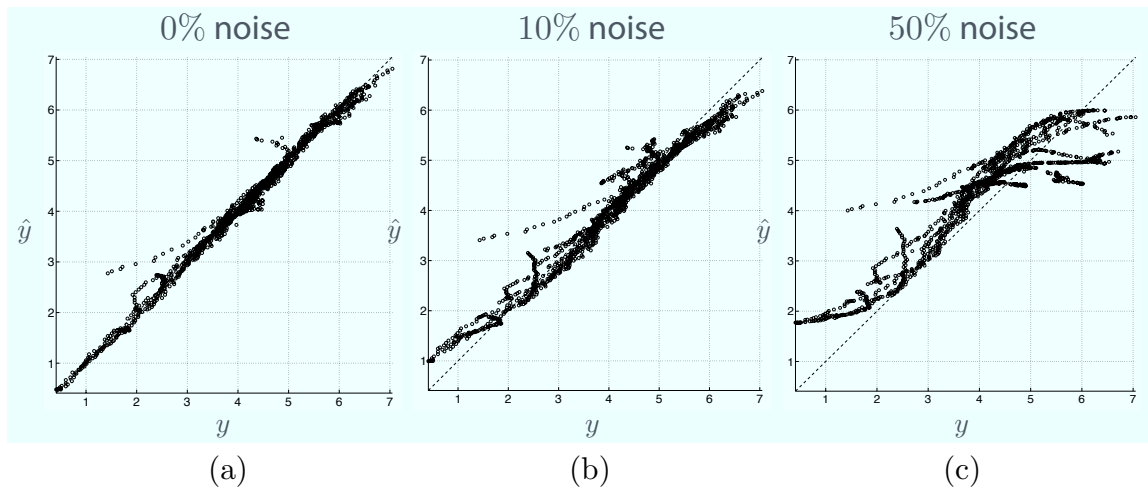


Figure 2.10: *Data-assimilation of coastal data: Left: Robustness of the riging predictor with respect to corrupted measurements. (Left to right) Predictions $\hat{y}$ versus corrupted data y for increasing noise amplitudes.*

2.6.3 Effect of low correlation

Our objective here is to highlight the importance of cross-correlations between outputs of different fidelity in obtaining an accurate recursive co-kriging predictor. In absence of such correlations, the proposed methodology is not expected to behave much better than fitting a kriging predictor through the observed points at the highest fidelity output. Here, we demonstrate this remark through a fabricated example in one input dimension, and in absence of measurement noise. To this end, consider the high-fidelity output to return realizations of a sine function in 1D

$$f_h(x) = \sin(6\pi x) \quad (2.12)$$

Then, let $f_l(x)$ be a low-fidelity model described as

$$f_l(x) = c \sin(6\pi x) + \cos(2\pi x) + 3e^{-x/2}, \quad (2.13)$$

where c is a parameter that controls the degree of correlation between $f_l(x)$ and $f_h(x)$. Obviously, values of c around 1 yield the strongest correlation, while values away from 1 lead to very weak correlations between $f_h(x)$ and $f_l(x)$.

Assume now that we have 6 equispaced observations of $f_h(x)$ in $[0, 1]$. In Fig. 2.11 we present the result of fitting an ordinary kriging predictor with a squared exponential kernel on this dataset. Evidently, the resolution of the observations is not adequate to resolve the variability in the exact solution, returning an inaccurate predictor with high variance at locations away from the observed points.

Next, we try to construct a more accurate predictor by employing additional in-

formation from a lower fidelity model $f_l(x)$ through recursive co-kriging. To this end, we have considered two cases that reflect the dependence of the proposed methodology on the intensity of spatial cross-correlations between different fidelity levels. In the first case, we take $c = 0.5$ in Eq. 2.13, and train a recursive co-kriging predictor on a set of 11 equispaced low-fidelity observations supplemented by 6 equispaced data-points of the high-fidelity model. The result is shown in Fig. 2.12(a), yielding a reasonably good fit to the exact solution, with a low associated uncertainty. The beneficial effect of injecting low-fidelity data is also indicated by the scaling factor ρ in the auto-regressive scheme of Eq. 3.55. Here, the MLE-based training scheme returned $\rho = 1.9247$, suggesting a strong correlation between $f_l(x)$ and $f_h(x)$. This is in contrast with the behavior demonstrated in Fig. 2.12(b), where the scaling parameter in the low-fidelity model of Eq. 2.13 is adjusted to $c = 0.4$. This choice leads to a weak spatial cross-correlation between $f_l(x)$ and $f_h(x)$, and as a result the information carried by the 11 equispaced observations of $f_l(x)$ has a negligible contribution in improving the accuracy of the co-kriging predictor. In fact, the resulting predictor nearly collapses to the estimate obtained using only the 6 high-fidelity observations in the kriging model of Fig. 2.11. The inadequacy of the low-fidelity model to provide useful information in this case is further confirmed by the scaling factor ρ , with MLE returning $\rho = 0.1683$ – a value more than 10 times less than the one predicted for $c = 0.5$.

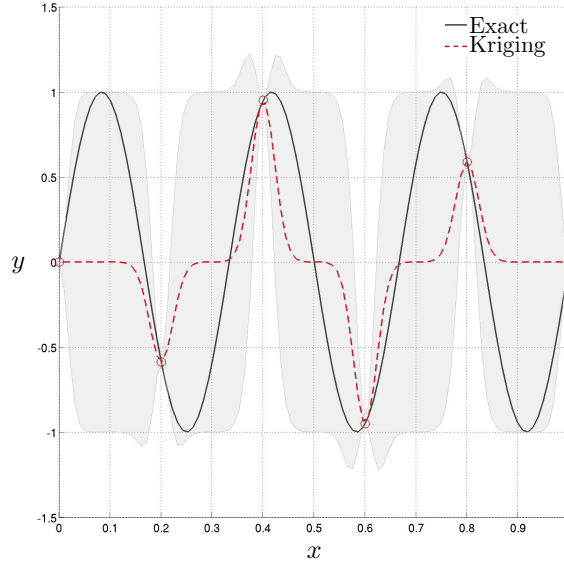


Figure 2.11: *Effect of low correlation:* Ordinary kriging predictor (*red dashed line*) through 6 equispaced observations (*red circles*) of $f_h(x)$ (*black solid line*). The gray shaded area corresponds to two standard deviations of the Gaussian predictor.

In summary, we have presented a case where injecting low-fidelity information leads to no improvement in the co-kriging predictor due to very weak spatial correlations between the low- and high-fidelity models. In such cases, one may expect that recursive co-kriging will result to nothing better than a kriging predictor through the highest fidelity points. However, the framework provides a mechanism for detecting this pathology, and that is done by monitoring the scaling factor ρ appearing in the auto-regressive scheme of Eq. 3.55, and is learned from the data via MLE.

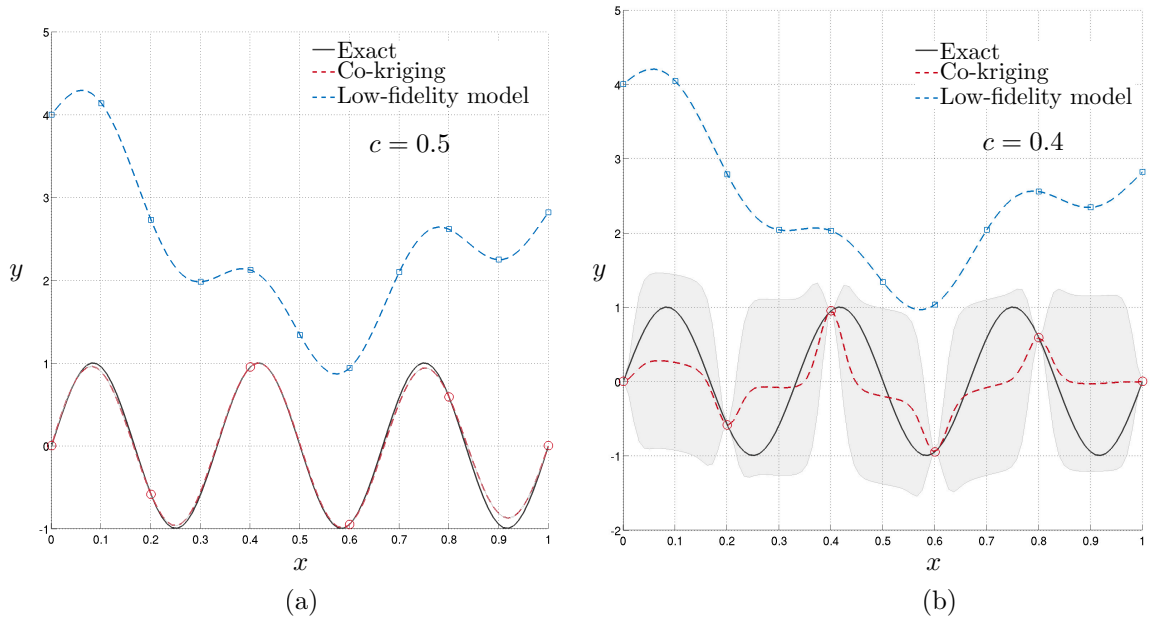


Figure 2.12: *Effect of low correlation: Left:* Two-level recursive co-kriging predictor for a low-fidelity model $f_l(x)$ with $c = 0.5$. *Right:* Two-level recursive co-kriging predictor for a low-fidelity model $f_l(x)$ with $c = 0.4$. In both cases, the training dataset consists of 11 equispaced observations of $f_l(x)$, and 6 equispaced observations of $f_h(x)$, while the gray shaded area depicts two standard deviations of the Gaussian predictor.

CHAPTER THREE

Spectral/ hp Element Methods for Inference with Gaussian Markov Random Fields

3.1 Motivation

In this chapter we study the applicability and performance of spectral/*hp* element methods for spatial statistics using Gauss Markov models. This work is motivated by the recent findings of [Lindgren et al. \(2011\)](#) that highlight the link between Gaussian fields and Gaussian Markov random fields through finite element discretization of stochastic partial differential equations. This approach has very promising implications as it circumvents the need to invert the full and ill-conditioned covariance matrices resulting from standard kriging techniques, while it enables the use of fast and scalable algorithms, thus opening the path to efficiently performing statistical inference on very big data sets. Here, we introduce the use of high-order nodal spectral/*hp* element methods and highlight all related implementation aspects. Moreover, we describe how this approach can be efficiently integrated into standard model fitting and statistical inference techniques, and utilized towards performing kriging predictions that satisfy Dirichlet and/or Neumann boundary conditions. The proposed algorithms are integrated in the general multi-fidelity framework proposed in [Chapter 2](#), and results are presented for several benchmark problems.

3.2 Gaussian Markov random fields and the SPDE approach

Let $Y(\mathbf{x})$ be an observation of a latent Gaussian field, $u(\mathbf{x})$, under a zero mean Gaussian measurement noise, $\epsilon(\mathbf{x})$, with variance σ^2 ,

$$Y(\mathbf{x}) = u(\mathbf{x}) + \epsilon(\mathbf{x}), \quad (3.1)$$

and let $\mu(\mathbf{x})$ and $\kappa(\mathbf{x}, \mathbf{x}')$ be the mean value function and the auto-correlation functions of $u(\mathbf{x})$, respectively. Additional assumptions on $\mu(\mathbf{x})$ may distinguish the above linear spatial prediction model into *simple* Kriging (if μ is assumed to be known), *ordinary* Kriging (if μ is unknown but independent of $\mathbf{x}$), and *universal* Kriging (if μ is unknown and is represented as a linear combination of deterministic basis functions, see [Cressie and Cassie \(1993\)](#)).

3.2.1 The Matérn covariance model

A popular choice for the covariance functions in spatial statistics is the Matérn covariance family (see [Murphy \(2012\)](#), [Cressie and Cassie \(1993\)](#)),

$$C(h) = \frac{2^{1-\nu}\sigma^2}{(4\pi)^{\frac{d}{2}}\Gamma(\nu + \frac{d}{2})\kappa^{2\nu}}(\kappa\|h\|)^\nu K_\nu(\kappa\|h\|), \quad \mathbf{h} \in \mathbb{R}^d, \quad (3.2)$$

where $h = \|\mathbf{x} - \mathbf{x}'\|_2$, ν determines the mean-square differentiability of the underlying process, κ is a scaling parameter related to the correlation length of the process, ρ , as $\rho = \sqrt{8\nu}/\kappa$, σ^2 is the marginal variance of the process, while $\Gamma(\cdot)$ and $K_\nu(\cdot)$ are the Euler gamma and modified Bessel function of the second kind, respectively. Using the asymptotic property

$$\left| K_\nu(t) - \frac{\Gamma(\nu)2^{\nu-1}}{t^\nu} \right| \rightarrow 0, \text{ as } t \rightarrow 0, \quad (3.3)$$

we can obtain a closed form expression for the variance of a process with Matérn covariance, as

$$C(0) = \frac{\pi^{\frac{d}{2}} \sigma^2 \Gamma(\nu)}{\Gamma(\nu + \frac{d}{2}) \kappa^{2\nu}} \quad (3.4)$$

Moreover, we can express the associated spectral density of the Matérn covariance function as

$$S(\mathbf{k}) = \frac{\sigma^2}{(\kappa^2 + \|\mathbf{k}\|^2)^{\nu + \frac{d}{2}}} \quad (3.5)$$

Note that for the spacial case of $\nu = 0.5$, the Matérn covariance simply reduces the exponential covariance model, while when $\nu \rightarrow \infty$ we recover the Gaussian covariance function (see [Murphy \(2012\)](#)).

3.2.2 Connection to stochastic partial differential equations (SPDE)

A powerful result by [Whittle \(1963\)](#) shows that a random process $u(\mathbf{x})$ with spectral density given by Eq. 3.5 is a solution to the SPDE

$$(\kappa^2 - \nabla^2)^{\frac{\alpha}{2}} u(\mathbf{x}) = \mathcal{W}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0, \quad (3.6)$$

where $\mathcal{W}(\mathbf{x})$ is Gaussian white noise. In literature, solutions to Eq. 3.6 are referred to as *Matérn fields* and are proved to be the only *stationary* solutions to this SPDE (see [Whittle \(1963\)](#)).

The key idea put forth by [Lindgren et al. \(2011\)](#) was to employ Hilbert space approximation theory to approximate the solution to Eq. 3.6 using a projection onto

a finite dimensional space spanned by basis functions $\phi_1(\mathbf{x}), \dots, \phi_n(\mathbf{x})$ as

$$u(\mathbf{x}) = \sum_{k=0}^n \phi_k(\mathbf{x}) w_k \quad (3.7)$$

where $\{w_k\}$ are a set of unknown expansion coefficients. In what follows, we will refer to the basis functions $\{\phi_k(\mathbf{x})\}$ using the terms “modes” and “basis functions” interchangeably, while the term “degrees of freedom” will refer to the expansion coefficients $\{w_k\}$.

A critical observation of [Lindgren et al. \(2011\)](#), citing the work of [Rozanov \(1982\)](#), states that a random field has a Markov property if and only if the reciprocal of its spectrum is a polynomial. Thereby, the Matérn fields generated by Eq. 3.6, having a spectrum given by Eq. 3.5, inherit the Markov property when α assumes integer values. Exploiting this Markov property, [Lindgren et al. \(2011\)](#) were able to utilize the finite dimensional representation of $u(\mathbf{x})$ given in Eq. 3.7 to construct a Gauss Markov random field (GMRF) with local neighborhood and precision matrix Q that *represents* the Gaussian field $u(\mathbf{x})$ in the sense that Q^{-1} is close to the covariance of $u(\mathbf{x})$, denoted by Σ , in some norm. The main advantage here is that the Markov property, along with the choice of a suitable basis, results to a sparse precision matrix Q , and, thus, enables the use of fast numerical methods for sparse matrices, significantly enhancing the computational efficiency of Bayesian inference.

3.3 Hilbert space approximation

3.3.1 Weak form

Consider the SPDE of Eq. 3.6 with $\alpha = 2$ in a two- or three-dimensional region Ω with boundary $\partial\Omega$:

$$(\kappa^2 - \nabla^2)u(\mathbf{x}) = \mathcal{W}(\mathbf{x}), \quad (3.8)$$

subject to

$$u|_{\partial\Omega^{\mathcal{D}}} = q_{\mathcal{D}}(\partial\Omega^{\mathcal{D}}), \quad \nabla u \cdot \mathbf{n}|_{\partial\Omega^{\mathcal{N}}} = q_{\mathcal{N}}(\partial\Omega^{\mathcal{N}}), \quad (3.9)$$

where $q_{\mathcal{D}}$ and $q_{\mathcal{N}}$ are Dirichlet and Neumann boundary conditions, $\mathbf{n}$ is the outward normal, and $\partial\Omega^{\mathcal{D}} \cup \partial\Omega^{\mathcal{N}} = \partial\Omega$. Introduction of the inner product $(f, g) = \int_{\Omega} f(\mathbf{x})g(\mathbf{x})d\mathbf{x}$ allows us to construct the weak form of this problem by multiplying Eq. 3.8 by a test function $v(\mathbf{x})$ that is homogeneous on all Dirichlet boundaries, that is, $v(\mathbf{x})$ belongs to the *test* space $\mathcal{V} = \{v \mid v \in H^1, v(0) = 0\}$. Integrating over the domain Ω we obtain

$$(\kappa^2 u, v) - (\nabla^2 u, v) = (\mathcal{W}, v) \quad (3.10)$$

Applying the divergence theorem to the diffusion term we have

$$(\kappa^2 u, v) + (\nabla u, \nabla v) = (\mathcal{W}, v) + \int_{\partial\Omega} v \nabla u \cdot \mathbf{n} dS, \quad (3.11)$$

Once a finite dimensional expansion basis for representing the solution $u(\mathbf{x})$ over a tessellation of the domain Ω is chosen, the discrete form of Eq. 3.11 reads as

$$\mathbf{H}\mathbf{w} = \mathbf{f} \quad (3.12)$$

where $\mathbf{w}$ is the vector of the unknown Galerkin coefficients of the solution expansion modes, and H is the discrete Helmholtz operator defined as

$$\mathbf{H} = \kappa^2 \mathbf{M} + \mathbf{S}, \quad (3.13)$$

with $\mathbf{M}$ and $\mathbf{S}$ being the mass and stiffness matrices, respectively, resulting from projecting the solution $u(\mathbf{x})$ onto a finite dimensional subspace spanned by the expansion modes $\{\phi_k(\mathbf{x})\}$, with entries computed as

$$M_{ij} = (\phi_i, \phi_j) \quad (3.14)$$

$$S_{ij} = (\nabla \phi_i, \nabla \phi_j) \quad (3.15)$$

Finally, the right hand side term $\mathbf{f} = \mathbf{f}_{\mathcal{W}} + \mathbf{f}_{BC}$ appearing in Eq. 3.12 results from the projection of the white noise forcing onto the basis, plus the forcing due to boundary fluxes, respectively,

$$\mathbf{f}_{\mathcal{W}} = (\mathcal{W}, v) \quad (3.16)$$

$$\mathbf{f}_{BC} = \int_{\partial\Omega} v \nabla u \cdot \mathbf{n} dS \quad (3.17)$$

3.3.2 Choosing an expansion basis

The choice of the expansion basis in Eq. 3.7 is of critical importance both in terms of obtaining an accurate solution to the SPDE, as well as for developing robust and computationally efficient Bayesian inference algorithms.

To illustrate the important role played by the choice of the expansion basis, let us start from computing the expectation of the Galerkin expansion coefficients $\mathbf{w}$ using Eq. 3.12

$$E[\mathbf{w}] = \mathbf{H}^{-1} \mathbf{f}_{BC} \quad (3.18)$$

Similarly, we can compute the covariance matrix of the Galerkin expansion coefficients, Σ_w , as

$$\Sigma_w = E[\mathbf{w}\mathbf{w}^T] - E[\mathbf{w}]E[\mathbf{w}^T] = \mathbf{H}^{-1} E[\mathbf{f}_{\mathcal{W}}\mathbf{f}_{\mathcal{W}}^T] \mathbf{H}^{-1} = \mathbf{H}^{-1} \mathbf{M} \mathbf{H}^{-1} \quad (3.19)$$

where we have used that $\mathcal{W}(\mathbf{x})$ is a zero mean Gaussian process, the forcing term $\mathbf{f}$ is a deterministic quantity, and the matrix $\mathbf{H}$ is symmetric. Therefore, the Galerkin expansion coefficients Σ_w form a Gaussian random field distributed as $\mathbf{w} \sim \mathcal{N}(\mathbf{H}^{-1} \mathbf{f}_{BC}, \Sigma_w)$. Consequently, the precision matrix of the Galerkin expansion coefficients, $\mathbf{Q}_w$, is computed as

$$\mathbf{Q}_w = \Sigma_w^{-1} = \mathbf{H} \mathbf{M}^{-1} \mathbf{H} \quad (3.20)$$

Then, the precision matrix of the solution field $u(\mathbf{x})$, denoted by $\mathbf{Q}_u$, can be recovered

as

$$\mathbf{Q}_u = \mathbf{\Phi} \mathbf{Q}_w \mathbf{\Phi}^T \quad (3.21)$$

where $\mathbf{\Phi}$ is a matrix containing the basis functions $\{\phi_k(\mathbf{x})\}$ evaluated at the grid point locations $\mathbf{x}$ that define the tessellation of Ω .

We can now advocate that the effect of choosing a suitable set of basis functions is threefold. First, the choice of basis determines the stability, converge and accuracy properties of the discretization scheme to Eq. 3.8. Second, it determines the structure, degree of sparsity and condition number of the mass, stiffness and Helmholtz operators, which, in turn, are responsible for passing these properties to the precision and covariance matrices of the Galerkin coefficients $\mathbf{w}$, according to Eqs. 3.19,3.20. Finally, the chosen basis affects the structure, degree of sparsity and condition number of the precision matrix, $\mathbf{Q}_u$, of the solution field, $u(\mathbf{x})$, as it determines the sparsity of the basis matrix $\mathbf{\Phi}$ in Eq. 3.21.

Lindgren et al. (2011) have addressed these issues by choosing to discretize Eq. 3.8 using a linear finite element basis. This choice is appealing as it immediately implies that the basis matrix $\mathbf{\Phi}$ in Eq. 3.21 is the identity matrix. Moreover, as a result of mass matrix lumping, the mass matrix is diagonal, suggesting that the finite dimensional approximation to the Gaussian random field $u(\mathbf{x})$ is a Gaussian Markov random field (GMRF) with sparse precision matrix. This consequence is highly desirable as it enables the use of fast algorithms for sparse matrices at the model fitting stage. However, we should note here that the mass matrix lumping technique employed by Lindgren et al. (2011) is an approximation that may degrade the accuracy and converge properties of the linear finite element basis (see Karniadakis and Sherwin (2013)), although results of Bolin and Lindgren (2009) have shown negligible differences in the resulting kriging errors between the exact finite element

representation (no mass lumping) and the Markov approximation (mass lumping).

[Bolin and Lindgren \(2009\)](#) introduced B-spline and Daubechies wavelets bases to demonstrate the advantages of the SPDE approach over two of the most popular covariance approximation techniques in spatial statistics, namely, covariance tapering and the convolution-fields method.

3.4 The spectral/*hp* element method

3.4.1 Overview of the method

Here, we advocate the use of high-order nodal spectral/*hp* element bases that attain all the aforementioned desired properties while providing a flexible framework for the development of fast and scalable algorithms. We begin with providing a brief overview of the discretization method and the numerical scheme employed, while for a complete treatment we refer to [Karniadakis and Sherwin \(2013\)](#).

First, the computational domain is formed by constructing a tessellation of Ω that may consist of triangular and/or quadrilateral elements in 2D, or a combination of tetrahedra, hexahedra, prisms, and pyramids in 3D. Within each of the N_{el} elements, the solution to the SPDE (Eq. 3.8) is approximated in terms of a polynomial expansion of the form

$$u(\mathbf{x}) = \sum_{i=1}^{N_{dof}} w_i \Phi_i(\mathbf{x}) = \sum_{e=1}^{N_{el}} \sum_{p=0}^P w_p^e \phi_p^e(\mathbf{x}_e(\xi)), \quad (3.22)$$

where N_{dof} is the total number of degrees of freedom, ξ defines a mapping from

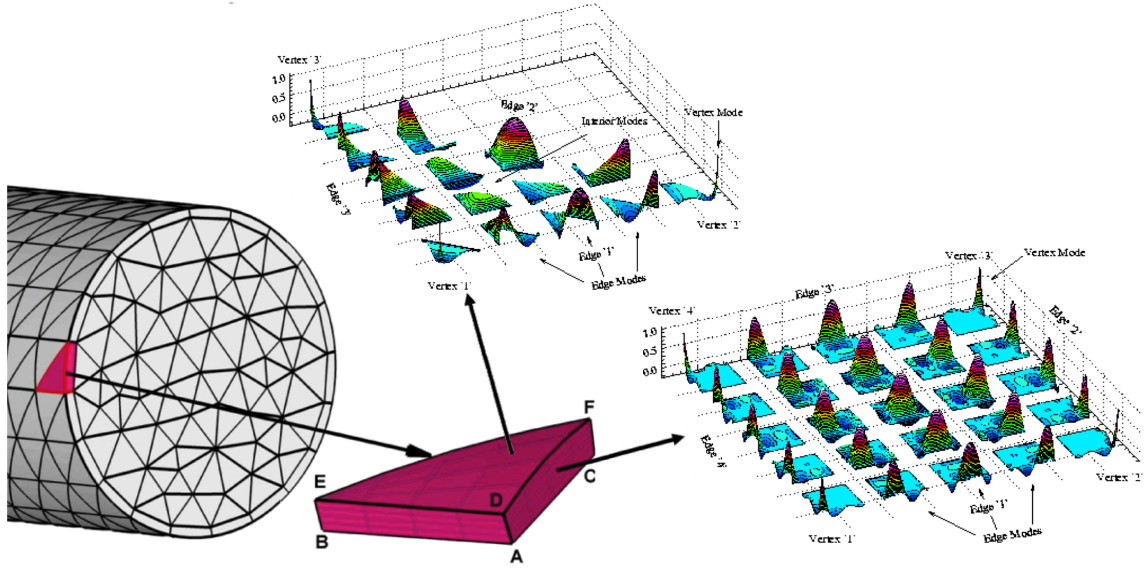


Figure 3.1: *The spectral/hp element method*: Sketch of the unstructured tessellation of the SPDE domain Ω the polynomial basis employed. The solution domain is decomposed into non-overlapping elements; within each element the solution is approximated by vertex, edge, face and (in 3D) interior modes. The shape functions associated with the vertex, edge and face modes for a fourth-order polynomial expansion ($P=4$) defined on triangular and quadrilateral faces are shown in color.

the physical space to the *standard* element, and $\phi_p^e(\mathbf{x}_e(\xi))$ are local to each element polynomials of order P , which when assembled together under the mapping $\mathbf{x}_e(\xi)$ result to a C^0 continuous global expansion $\Phi_p(\mathbf{x})$. The modes can be distinguished between vertex (linear) modes, edge, face, and bubble (interior) modes as depicted in Fig. 3.1. Note that the interior modes have zero support on the elemental boundaries, thus the boundary and interior degrees of freedom can be numerically decoupled through a technique known as *substructuring*, where the Schur complement of the boundary system is constructed. The boundary degrees of freedom, corresponding to adjacent elements, are coupled due to the requirement of C^0 -continuity.

The local p -type basis functions employed in this study are constructed using the nodal Legendre cardinal functions, yielding the following expression in the standard element

$$\phi_p^e(\xi) = \frac{(\xi - 1)(\xi + 1)\partial_\xi L_p(\xi)}{P(P + 1)L_p(\xi_P)(\xi - \xi_P)}, \quad 0 \leq p \leq P, \quad (3.23)$$

where L_P is the Legendre polynomial of order P . This nodal p -type basis has become known in the literature as “spectral elements”, due to [Patera \(1984\)](#), although in the original work of Patera, Chebyshev polynomials were used instead of Legendre.

Two- and three-dimensional basis expansions can be constructed in terms of a tensor product of the above one-dimensional expansion as:

$$\phi_{pq}^e(\xi_1, \xi_2) = \phi_p^e(\xi_1)\phi_q^e(\xi_2), \quad 0 \leq p, q; \quad p \leq P_1, \quad q \leq P_2 \quad \text{in 2D} \quad (3.24)$$

$$\phi_{pqr}^e(\xi_1, \xi_2, \xi_3) = \phi_p^e(\xi_1)\phi_q^e(\xi_2)\phi_r^e(\xi_3), \quad 0 \leq p, q, r; \quad p \leq P_1, \quad q \leq P_2, \quad r \leq P_3 \quad \text{in 3D}, \quad (3.25)$$

where the polynomial order of the multi-dimensional expansions may differ in each coordinate direction as denoted by the bounds P_1, P_2, P_3 . Figure [3.1](#) illustrates the resulting basis defined on triangular and quadrilateral faces for $P = 4$.

3.4.2 Properties of the nodal basis

The choice of this nodal spectral element basis has several advantageous properties. First, it results to a sparse, symmetric positive-definite stiffness matrix, $\mathbf{S}$. Moreover, under-integration of the inner products defining the entries of the mass matrix, $\mathbf{M}$, using the Gauss-Legendre-Lobatto quadrature at the $(P + 1)$ nodal locations, causes $\mathbf{M}$ to become diagonal due to the discrete orthogonality property of the basis (see [Karniadakis and Sherwin \(2013\)](#)). The error introduced from this reduced order

quadrature rule is, however, consistent with the approximation error of the expansion (see [Karniadakis and Sherwin \(2013\)](#)). In contrast with the mass matrix lumping technique employed in the finite element formulation of [Lindgren et al. \(2011\)](#), the inexact integration employed in the spectral element setting is not an approximation, and has a direct equivalence to the error of approximating a smooth function (see [Karniadakis and Sherwin \(2013\)](#)).

Another favorable property of this basis arises when the numerical quadrature order is equal to the polynomial order plus one (same as above). Then, the modal expansion coefficients are simply the physical values of the solution at the quadrature points, and the basis matrix Φ appearing in Eq. 3.21 simply reduces to the identity matrix. This, along with the diagonal property of the mass matrix, naturally implies that the approximation to the solution of the SPDE in Eq. 3.8 will be a Gaussian Markov random field (GMRF) with a sparse precision matrix, given by Eq. 3.21.

3.4.3 Treatment of boundary conditions

In order to enforce the set of boundary conditions described in Eq. 3.9 we recall that the space of the test functions $v(\mathbf{x})$ in a Galerkin expansion is homogeneous and is therefore zero on all Dirichlet boundaries. As we can only solve for the components of $u(\mathbf{x})$ that live in the same space, we employ the technique of *lifting* a known solution (see [Karniadakis and Sherwin \(2013\)](#)) and decompose the solution $u(\mathbf{x})$ into its homogeneous $u^{\mathcal{H}}(\mathbf{x})$ and non-homogeneous $u^{\mathcal{D}}(\mathbf{x})$ components, where $u(\mathbf{x}) = u^{\mathcal{H}}(\mathbf{x}) + u^{\mathcal{D}}(\mathbf{x})$, and $u^{\mathcal{D}}(\mathbf{x})$ satisfies all Dirichlet boundary conditions $q_{\mathcal{D}}(\partial\Omega^{\mathcal{D}})$. We note that the important component of this definition is that the homogeneous solution has zero contribution from modes which are nonzero on Dirichlet boundaries. The homogeneous solution also contains the interior degrees of freedom which are

defined to be zero on the boundaries. Moreover, recalling that $v(\mathbf{x})$ is defined to vanish on all Dirichlet boundaries we can state the Galerkin problem of Eq. 3.11 as: Find $u^{\mathcal{H}}$ in the *trial* space $\mathcal{X} = \{u \mid u \in H^1, u(0) = q_{\mathcal{D}}\}$ such that

$$(\kappa^2 u^{\mathcal{H}}, v) + (\nabla u^{\mathcal{H}}, \nabla v) = (\mathcal{W}, v) + \int_{\partial\Omega^{\mathcal{H}}} v q_N dS - (\nabla u^{\mathcal{D}}, \nabla v) \quad (3.26)$$

3.4.3.1 Solving the global matrix system

Recalling that the boundary and interior degrees of freedom are decoupled, we can order the global boundary degrees of freedom (indexed by the superscript b) in a continuous fashion, such that degrees of freedom associated with the unknown homogeneous part of the solution $u^{\mathcal{H}}(\mathbf{x})$ are followed by the known degrees of freedom associated with the non-homogenous part $u^{\mathcal{D}}(\mathbf{x})$ that is defined through Dirichlet boundary conditions. Finally, the decoupled interior degrees of freedom (indexed by the superscript i) are placed at the end of the pile. Following this numbering scheme, the global matrix system introduced in Eq. 3.12 takes the form

$$\underbrace{\begin{bmatrix} H_{\mathcal{H}\mathcal{H}} & H_{\mathcal{H}\mathcal{D}} & H_{\mathcal{H}I} \\ H_{\mathcal{D}\mathcal{H}} & H_{\mathcal{D}\mathcal{D}} & H_{\mathcal{D}I} \\ \hline H_{I\mathcal{H}} & H_{I\mathcal{D}} & H_{II} \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} w_{\mathcal{H}}^b \\ w_{\mathcal{D}}^b \\ \hline w_{\mathcal{H}}^i \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} f_{\mathcal{H}}^b \\ f_{\mathcal{D}}^b \\ \hline f_{\mathcal{H}}^i \end{bmatrix}}_{\mathbf{f}} \quad (3.27)$$

where the vertical and horizontal lines are used to distinguish between the boundary and interior contributions.

Using the technique known as *substructuring*, we can statically condense the global matrix system by removing the known, non-homogeneous boundary degrees

of freedom that correspond to Dirichlet boundary conditions, $w_{\mathcal{D}}^b$,

$$\underbrace{\begin{bmatrix} H_{\mathcal{H}\mathcal{H}} & H_{\mathcal{H}I} \\ H_{I\mathcal{H}} & H_{II} \end{bmatrix}}_{\mathbf{H}_{\mathcal{H}}} \underbrace{\begin{bmatrix} w_{\mathcal{H}}^b \\ w_{\mathcal{H}}^i \end{bmatrix}}_{\mathbf{w}_{\mathcal{H}}} = \underbrace{\begin{bmatrix} f_{\mathcal{H}}^b - H_{\mathcal{H}\mathcal{D}}w_{\mathcal{D}}^b \\ f_{\mathcal{H}}^i - H_{I\mathcal{D}}w_{\mathcal{D}}^b \end{bmatrix}}_{\mathbf{f}_{\mathcal{H}}}, \quad (3.28)$$

Then, we can separately solve for the unknown boundary and interior degrees of freedom in two steps as,

$$S_{\mathcal{H}\mathcal{H}}w_{\mathcal{H}}^b = f_{\mathcal{H}}^b - H_{\mathcal{H}I}H_{II}^{-1}f^i - S_{\mathcal{H}\mathcal{D}}w_{\mathcal{D}}^b \quad (3.29)$$

$$H_{II}w_{\mathcal{H}}^i = f^i - H_{I\mathcal{H}}w_{\mathcal{H}}^b - H_{I\mathcal{D}}w_{\mathcal{D}}^b, \quad (3.30)$$

where the matrix S is known as the Schur complement and is computed as,

$$S = \begin{bmatrix} S_{\mathcal{H}\mathcal{H}} & S_{\mathcal{H}\mathcal{D}} \\ S_{\mathcal{D}\mathcal{H}} & S_{\mathcal{D}\mathcal{D}} \end{bmatrix} = H_{\mathcal{H}\mathcal{D}} - H_{\mathcal{H}I}H_{II}^{-1}H_{I\mathcal{D}} \quad (3.31)$$

This solution framework gives rise to a number of observations. First, we note that the size of the Schur complement matrix $S_{\mathcal{H}\mathcal{H}}$ in Eq. 3.31 can be significantly smaller than the global Helmholtz matrix $\mathbf{H}$, while its condition number can be no larger than the condition number of the complete system in Eqs. 3.12,3.27 (see Karniadakis and Sherwin (2013)). Moreover, the Schur complement matrix $S_{\mathcal{H}\mathcal{H}}$ is symmetric and positive definite, allowing for the use of efficient and scalable preconditioners to accelerate the iterative solution of Eq. 3.29 (see Sherwin and Casarin (2001), Grinberg et al. (2009)). Second, since the interior modes are decoupled between elements, Eq. 3.30 can be solved in parallel, locally to each element. If an iter-

ative solver is used for Eq. 3.29, the method requires the storage only of the following matrices, and only at an elemental level (no global matrix assembly is required (see Karniadakis and Sherwin (2013))): $S, H_{\mathcal{H}I}H_{II}^{-1}, H_{I\mathcal{H}}, H_{ID}, H_{II}^{-1}$. In fact, the square matrices S, H_{II}^{-1} are symmetric, hence their storage cost is further reduced. These features are of significant importance, as they provide a cornerstone for achieving scalability and computational efficiency, especially for very large systems.

3.4.4 Computation of the covariance and precision matrices

Recall the Galerkin formulation in Eq. 3.26, and let us compute the expected value of the expansion coefficients corresponding to the homogeneous part of the solution $u^{\mathcal{H}}(\mathbf{x})$

$$E[\mathbf{w}_{\mathcal{H}}] = \mathbf{H}_{\mathcal{H}}^{-1} \mathbf{f}_{BC}^{\mathcal{H}}, \quad (3.32)$$

where $\mathbf{H}_{\mathcal{H}}$ is the homogeneous Helmholtz operator, constructed from the blocks of the full Helmholtz system that correspond to the homogeneous modes, as shown in Eq. 3.52. Moreover, the deterministic forcing term $\mathbf{f}_{BC}^{\mathcal{H}}$ now accounts for both Dirichlet and/or Neumann boundary conditions, taking the form

$$\mathbf{f}_{BC}^{\mathcal{H}} = \int_{\partial\Omega^{\mathcal{H}}} v q_N dS - (\nabla u^{\mathcal{D}}, \nabla v) \quad (3.33)$$

Similarly, we can compute the covariance matrix of the homogeneous Galerkin expansion coefficients, $\Sigma_w^{\mathcal{H}}$, as

$$\Sigma_w^{\mathcal{H}} = E[\mathbf{w}_{\mathcal{H}} \mathbf{w}_{\mathcal{H}}^T] - E[\mathbf{w}_{\mathcal{H}}] E[\mathbf{w}_{\mathcal{H}}^T] = \mathbf{H}_{\mathcal{H}}^{-1} E[\mathbf{f}_{\mathcal{W}} \mathbf{f}_{\mathcal{W}}^T] \mathbf{H}_{\mathcal{H}}^{-1} = \mathbf{H}_{\mathcal{H}}^{-1} \mathbf{M}_{\mathcal{H}} \mathbf{H}_{\mathcal{H}}^{-1} \quad (3.34)$$

where we have used that $\mathcal{W}(\mathbf{x})$ is a zero mean Gaussian process, the forcing term $\mathbf{f}_{BC}^{\mathcal{H}}$ is a deterministic quantity, and the matrix $\mathbf{H}_{\mathcal{H}}$ is symmetric. Therefore, the Galerkin expansion coefficients, $\mathbf{w}_{\mathcal{H}}$, corresponding to the homogeneous part of the solution, $u^{\mathcal{H}}(\mathbf{x})$, form a Gaussian random field distributed as $\mathbf{w}_{\mathcal{H}} \sim \mathcal{N}(\mathbf{H}_{\mathcal{H}}^{-1} \mathbf{f}_{BC}^{\mathcal{H}}, \Sigma_w^{\mathcal{H}})$.

Consequently, the corresponding precision matrices $\mathbf{Q}_w^{\mathcal{H}}$, and $\mathbf{Q}_u^{\mathcal{H}}$ are computed as

$$\mathbf{Q}_w^{\mathcal{H}} = (\Sigma_w^{\mathcal{H}})^{-1} = \mathbf{H}_{\mathcal{H}} \mathbf{M}_{\mathcal{H}}^{-1} \mathbf{H}_{\mathcal{H}} = \mathbf{Q}_u^{\mathcal{H}} \quad (3.35)$$

since the nodal basis we have employed results to an identity matrix Φ in Eq. 3.21. Recall that the nodal spectral element basis with inexact integration results to a fully diagonal mass matrix $\mathbf{M}_{\mathcal{H}}$, therefore the homogeneous part of the solution $u^{\mathcal{H}}(\mathbf{x})$ to the SPDE (Eq. 3.8) has a sparse precision matrix, $\mathbf{Q}_u^{\mathcal{H}}$, hence is a Gaussian Markov random field. In fact, any other choice of basis that does not produce a fully diagonal mass matrix would not attain this property, jeopardizing computational efficiency in the model fitting process. To clarify this point, let n be the total number of homogeneous degrees of freedom of the nodal spectral element expansion, that is the matrices $\mathbf{H}_{\mathcal{H}}, \mathbf{M}_{\mathcal{H}}, \Sigma_w^{\mathcal{H}}, \mathbf{Q}_u^{\mathcal{H}}$ appearing in Eqs. 3.34, 3.35 are all square, symmetric positive definite, of size $(n \times n)$. It is clear that when $\mathbf{M}_{\mathcal{H}}$ is fully diagonal, the precision matrix $\mathbf{Q}_u^{\mathcal{H}}$ can be computed from Eq. 3.35 just at the cost of matrix multiplication, as the inversion of $\mathbf{M}_{\mathcal{H}}$ is straightforward. On the contrary, the computation of the covariance $\Sigma_w^{\mathcal{H}}$ in Eq. 3.34 requires an additional cost of up to $\mathcal{O}(n^3)$ operations due to the inversion of the Helmholtz operator $\mathbf{H}_{\mathcal{H}}$. Moreover, the resulting covariance $\Sigma_w^{\mathcal{H}}$ is a full matrix, as opposed to the precision $\mathbf{Q}_u^{\mathcal{H}}$ that

is sparse. Therefore, gains in computational efficiency are made if the model fitting process can be formulated in terms of operations involving the sparse precision matrix of the GMRF $u^{\mathcal{H}}(\mathbf{x}) \sim \mathcal{N}(\mathbf{H}_{\mathcal{H}}^{-1} \mathbf{f}_{BC}^{\mathcal{H}}, (\mathbf{Q}_u^{\mathcal{H}})^{-1})$. Finally, we note that $\mathbf{Q}_u^{\mathcal{H}}$ can be used to consistently sample the a GMRF that approximates the homogeneous part of the solution to Eq. 3.8. This solution vector can then be augmented with the known, non-homogeneous contribution arising from Dirichlet data to obtain a Matérn field that satisfies the SPDE (Eq. 3.8), subject to Dirichlet and/or Neumann boundary conditions (Eq. 3.9).

The most widely used way of efficiently accessing the covariance in practice is through the Cholesky decomposition of the precision, leveraging on its sparsity for the GMRF case (see Rue and Held (2005)). Here, we present an alternative way to construct the covariance matrix by explicitly computing each of its constituting blocks

$$\Sigma_w^{\mathcal{H}} = \begin{bmatrix} \Sigma_{\mathcal{H}\mathcal{H}} & \Sigma_{\mathcal{H}I} \\ \Sigma_{I\mathcal{H}} & \Sigma_{II} \end{bmatrix} = \Sigma_u^{\mathcal{H}}, \quad (3.36)$$

where each block reduces to the following expressions using Eqs. 3.29,3.30 and thanks to the properties of the nodal spectral element basis

$$\Sigma_{\mathcal{H}\mathcal{H}} = S_{\mathcal{H}\mathcal{H}}^{-1} M_{\mathcal{H}\mathcal{H}} (S_{\mathcal{H}\mathcal{H}}^{-1})^T + (S_{\mathcal{H}\mathcal{H}}^{-1} H_{\mathcal{H}I} H_{II}^{-1}) M_{II} (S_{\mathcal{H}\mathcal{H}}^{-1} H_{\mathcal{H}I} H_{II}^{-1})^T \quad (3.37)$$

$$\Sigma_{\mathcal{H}I} = (S_{\mathcal{H}\mathcal{H}}^{-1} H_{\mathcal{H}I} H_{II}^{-1}) M_{II} (H_{II}^{-1})^T \quad (3.38)$$

$$\Sigma_{I\mathcal{H}} = \Sigma_{\mathcal{H}I}^T = H_{II}^{-1} M_{II} (S_{\mathcal{H}\mathcal{H}}^{-1} H_{\mathcal{H}I} H_{II}^{-1})^T \quad (3.39)$$

$$\Sigma_{II} = H_{II} M_{II} (H_{II}^{-1})^T \quad (3.40)$$

Note that this approach requires only the inversion of the homogeneous-homogeneous block of the Schur complement matrix $S_{\mathcal{H}\mathcal{H}}$ and the interior-interior block of the

Helmholtz operator H_{II} . The first, can be significantly smaller in size than the global Helmholtz system in Eq. 3.34 and can be effectively preconditioned when solved iteratively, while the latter consists of fully decoupled blocks that can be inverted element-wise in parallel. Moreover, it requires no additional storage since the required matrices $S_{\mathcal{H}\mathcal{H}}, H_{\mathcal{H}I}H_{II}^{-1}, H_{II}^{-1}$ are already available from Eqs. 3.29, 3.30, 3.31.

Whether this approach results to computational gains versus the popular Cholesky decomposition will be the subject of a future study. Here, we note that this comparison is expected to be heavily influenced by the correlation length of the underlying process $u(\mathbf{x})$ and the resulting condition number of the global matrix system. With that in mind, the proposed approach may outperform the Cholesky decomposition for large systems and short correlation lengths.

3.5 Model fitting and the kriging predictor

In this section we provide a brief overview of the model fitting procedure needed to train the observation model of Eq. 3.1 on a set of observed data. For more details on parameter estimation for Bayesian inference we refer the reader to (see Murphy (2012), Rasmussen (2006), Rue and Held (2005)).

Generally speaking, a Bayesian hierarchical model typical consists of (at least) a *data model*, characterized by the conditional probability distribution $\pi(y|u)$ that describes how observations arise assuming a known process $u(\mathbf{x})$, a *latent model*, characterized by a conditional probability $\pi(u|\theta)$ describing how the latent process $u(\mathbf{x})$ behaves assuming a set of known parameters θ , and a set of *parameters* characterized by their probability distribution $\pi(\theta)$ which describes our, sometimes vague,

prior knowledge of the parameters.

Using Bayes' theorem, these probability distributions can be utilized to provide an expression for the *posterior* distribution $\pi(u|y)$ that characterizes the statistics of the latent process $u(\mathbf{x})$, given a set of observations y

$$\pi(u|y) = \frac{\pi(y|u)\pi(u)}{\pi(y)} \propto \pi(y|u)\pi(u) \quad (3.41)$$

Given arbitrary data we typically don't know the expectation or precision, but let these follow some known parametric structure that depends on a few parameters θ . Then,

$$\pi(y|u) \sim \mathcal{N}(Au, \mathbf{Q}_\epsilon^{-1}(\theta)), \quad \pi(u) \sim \mathcal{N}(\mu(\theta), \mathbf{Q}_u^{-1}(\theta)), \quad \pi(u|y) \sim \mathcal{N}(\mu_{u|y}(\theta), \mathbf{Q}_{u|y}^{-1}(\theta)) \quad (3.42)$$

where A is an operator that restricts the latent process $u(\mathbf{x})$ to the locations of the observed data, $\pi(u|y)$ is the distribution describing how the latent field $u(\mathbf{x})$ behaves given the some observations y , $\mathbf{Q}_\epsilon$ is the precision of the measurement noise appearing in the right hand side of our observation model, while $\mu(\theta)$ is the mean of the latent process $u(\mathbf{x})$ (see Eq. 3.1). At locations where observations exist, or where $u(\mathbf{x})$ satisfies a known set of Dirichlet boundary conditions, $\mu(\theta)$ attains these values up to measurement error, controlled by the zero-mean process $\epsilon(\mathbf{x})$ (see Eq. 3.1).

The model fitting process consists of estimating the parameters θ , and possibly evaluating the parameter uncertainty based on the observed data. To this end, we need to compute the following posterior

$$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta) = \int \pi(y, u|\theta)\pi(\theta)du = \int \pi(y|u, \theta)\pi(u|\theta)du \pi(\theta) \quad (3.43)$$

Instead of explicitly integrating over u , a simple trick yields a different derivation of $\pi(\theta|y)$:

$$\pi(\theta|y) \propto \pi(\theta)\pi(y|\theta) = \pi(\theta)\pi(y|\theta)\frac{\pi(z|y, \theta)}{\pi(z|y, \theta)}, \quad (3.44)$$

for any arbitrary z . Hence, we may pick $z = \mu_{u|y}$, and using the fact that $u(\mathbf{x})$ and $\epsilon(\mathbf{x})$ are Gaussian processes, we can express the posterior $\pi(y|\theta)$ as

$$\pi(y|\theta) \propto \frac{|\mathbf{Q}_u|^{1/2}|\mathbf{Q}_\epsilon|^{1/2}}{|\mathbf{Q}_{u|y}|^{1/2}} \exp\left(-\frac{1}{2}[(\mu_{u|y} - \mu)^T \mathbf{Q}_u(\mu_{u|y} - \mu) + (y - A\mu_{u|y})^T \mathbf{Q}_\epsilon(y - A\mu_{u|y})]\right), \quad (3.45)$$

where we recall that μ is the mean of the latent process $u(\mathbf{x})$. Then, the set of variables that parametrize the observation model of Eq. 3.1 can be estimated by maximizing the posterior of Eq. 3.45, as

$$\hat{\theta} = \arg \max_{\theta} \pi(y|\theta) \quad (3.46)$$

Moreover, the precision $\mathbf{Q}_{u|y}$, mean $\mu_{u|y}$, and variance $\text{Var}(u|y)$ can be explicitly expressed as

$$\mathbf{Q}_{u|y} = \mathbf{Q}_u + A^T \mathbf{Q}_\epsilon A \quad (3.47)$$

$$\mu_{u|y} = \mu + \mathbf{Q}_{u|y}^{-1} A^T \mathbf{Q}_\epsilon (y - A\mu) \quad (3.48)$$

$$\text{Var}(u|y) = \mathbf{Q}_{u|y}^{-1} \quad (3.49)$$

Once the optimization problem for identifying the model parameters is solved (Eq. 3.46), the kriging predictor and kriging variance of the observation model (Eq. 3.1) are readily available through $\mu_{u|y}$, and $\text{Var}(u|y)$ (Eqs. 3.48, 3.49), respectively.

The step that determines the efficiency of the parameter estimation workflow is the inversion of the precision $\mathbf{Q}_{u|y}^{-1}$ in computing $\mu_{u|y}$ (Eq. 3.48). The most widely used approach is based on the Cholesky decomposition of the precision $\mathbf{Q}_u$ and can be outlined in the following steps (see Rue and Held (2005)):

1. Rewrite Eq. 3.48 as: $\mu_{u|y} = \mu - \mathbf{Q}_u^{-1} A^T (A \mathbf{Q}_u^{-1} A^T + \mathbf{Q}_\epsilon^{-1})^{-1} (A\mu - y)$
2. Compute the Cholesky decomposition of $\mathbf{Q}_u$ such that: $\mathbf{Q}_u = R^T R$
3. Use back-substitution to calculate $B = \mathbf{Q}_u^{-1} A^T = R^{-1} (R^T)^{-1} A^T$
4. Compute R_2 using the Cholesky decomposition of $AB + \mathbf{Q}_\epsilon^{-1} = R_2^T R_2$
5. Compute $C = B(AB + \mathbf{Q}_\epsilon^{-1})^{-1} = B R_2^{-1} (R_2^T)^{-1}$
6. Compute $\mu_{u|y} = \mu - C(A\mu - y)$

Alternatively, if we employ the block covariance method (Eq. 3.36) proposed in the last part of Sec. 3.4.4, we can compute $\mu_{u|y}$ as:

1. Rewrite Eq. 3.48 as: $\mu_{u|y} = \mu - \Sigma_u A^T (A \Sigma_u A^T + \mathbf{Q}_\epsilon^{-1})^{-1} (A\mu - y)$

2. Compute $B = \Sigma_u A^T$
3. Compute the Cholesky decomposition of $AB + \mathbf{Q}_\epsilon^{-1} = R_2^T R_2$
4. Compute $C = B(AB + \mathbf{Q}_\epsilon^{-1})^{-1} = BR_2^{-1}(R_2^T)^{-1}$
5. Compute $\mu_{u|y} = \mu - C(A\mu - y)$

where in both cases, the size of the matrix $(AB + \mathbf{Q}_\epsilon^{-1})$ is determined by the number of observations, hence it is typically expected to be much smaller than the size $\mathbf{Q}_u$ or Σ_u themselves.

3.5.1 Maximum Likelihood Estimation with GMRFs

In practice, the optimization problem for estimating the model parameters θ (Eq. 3.46) is formulated as a global minimization problem of the negative log-likelihood $\mathcal{L}$:

$$\mathcal{L} = -\frac{1}{2} \left[\log \mathbf{Q}_u + \log \mathbf{Q}_\epsilon - \log \mathbf{Q}_{\mathbf{u}|\mathbf{y}} - (\mu_{u|y} - \mu)^T \mathbf{Q}_u (\mu_{u|y} - \mu) - (y - A\mu_{u|y})^T \mathbf{Q}_\epsilon (y - A\mu_{u|y}) \right], \quad (3.50)$$

where recall that all $\mathbf{Q}_{u|y}$, $\mathbf{Q}_u$, $\mathbf{Q}_\epsilon$, $\mu_{u|y}$, μ are parametrized by the corresponding entry in the θ vector (see Eq. 3.42). Minimization of $\mathcal{L}$ is typically achieved through gradient descent algorithms. Once the solver has converged to a minimum and a set of optimal parameters $\hat{\theta}$ has been obtained, it is straightforward to perform kriging predictions and estimate the kriging variance through $\mu_{u|y}$, and $\text{Var}(u|y)$ (Eqs. 3.48, 3.49), respectively.

3.6 Multi-fidelity modeling via recursive co-kriging and Gaussian Markov random fields

The proposed algorithms can be integrated in the general multi-fidelity information fusion framework introduced in Chapter 2. Here, we outline the procedure of recasting the problem in terms of GMFRs, and provide an overview of the key implementation aspects.

3.6.1 Ordinary kriging with GMRFs

Here we present a brief overview of the kriging predictive scheme leveraging the sparsity of GMRF precision matrices. The reader is referred to [Lindgren et al. \(2011\)](#), [Rasmussen \(2006\)](#) for a detailed exposition of the subject. As usual, we model scattered observations of a field $Y(\mathbf{x})$ as a realization of a Gaussian random field $Z(\mathbf{x})$. These observations could be corrupted by model error or measurement noise that is described by a Gaussian process $\mathcal{E}(\mathbf{x})$, leading to an observation model of the form

$$Y(\mathbf{x}) = Z(\mathbf{x}) + \mathcal{E}(\mathbf{x}). \quad (3.51)$$

Consequently, the random field $(Z(\mathbf{x}), Y(\mathbf{x}))^T$ has a multivariate normal distribution

$$\begin{bmatrix} Z(\mathbf{x}) \\ Y(\mathbf{x}) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu \\ A\mu \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma A^T \\ A^T \Sigma & A \Sigma A^T + \Sigma_{\mathcal{E}} \end{bmatrix} \right), \quad (3.52)$$

where, $\mu = E[Z]$ is the mean value of Z , Σ is the covariance of Z , $\Sigma_{\mathcal{E}}$ is the covariance of the noise $\mathcal{E}(\mathbf{x})$, and A is a simple matrix that restricts the random field $Z(\mathbf{x})$ to the locations of observed data y .

Starting from the conditional distribution of $Z|Y$ we can derive equations for the expectation $\mu_{Z|Y}(\mathbf{x})$, and covariance $\Sigma_{Z|Y}(\mathbf{x})$ of the ordinary kriging predictor as

$$\mu_{Z|Y}(\mathbf{x}) = \mu + (\Sigma A^T)(A \Sigma A^T + \Sigma_{\mathcal{E}})^{-1}(y - A\mu), \quad (3.53)$$

$$\Sigma_{Z|Y}(\mathbf{x}) = \Sigma - (\Sigma A^T)(A \Sigma A^T + \Sigma_{\mathcal{E}})^{-1}(A^T \Sigma). \quad (3.54)$$

Typically, the covariances Σ and $\Sigma_{\mathcal{E}}$ are parametrized by a set of hyper-parameters, $\{\theta, \theta_{\mathcal{E}}\}$, that can be *learned* from the data y by maximizing the posterior $\pi(y|\theta, \theta_{\mathcal{E}})$ using Maximum Likelihood Estimation (MLE) or in a fully Bayesian setting (e.g. Markov Chain Monte Carlo integration). This essentially introduces a machine learning problem for computing the optimal values of the unknown parameters $\{\mu, \theta, \theta_{\mathcal{E}}\}$ from the data.

The noise process $\mathcal{E}(\mathbf{x})$ accounts for *epistemic* uncertainties that lead to differences between our observables and the real response of the system due to inaccurate or uncertain modeling assumptions. In the experimental setting this may correspond to noise or bias in the measurements, while in the numerical setting it may account for round-off or truncation errors of the numerical scheme. In general, modeling the nature of $\mathcal{E}(\mathbf{x})$ results in an intrinsically hard problem that is beyond the scope of the present study. Without loss of generality, in what follows we will assume that $\mathcal{E}(\mathbf{x})$ is a zero-mean Gaussian white noise process with variance $\sigma_{\mathcal{E}}$, i.e $\mathcal{E}(\mathbf{x}) \sim \mathcal{N}(0, \sigma_{\mathcal{E}} I)$. This should not be confused with *aleatoric* uncertainties that arise from randomness in the input parameters of the system that will be taken into account by the random vector ξ (see Fig. 2.1, in Chapter 2, Sec. 2.2).

3.6.1.1 Ordinary co-kriging with GMRFs

The idea of kriging is naturally extended to the multivariate setting, in which the observation model for $Y(\mathbf{x})$ depends on more than one covariates. Here, we present a brief overview of the approach of [Kennedy and O'Hagan \(2000\)](#), which introduced a co-kriging model based on a first-order autoregressive relation between model output of different levels of fidelity.

Suppose we have s levels of variable-fidelity model output $(y_t(\mathbf{x}))_{t=1}^s$ at locations $\mathbf{x} \in D_{t=1}^s$ sorted by increasing fidelity and modeled as observations of a Gaussian field $(Z_t(\mathbf{x}))_{t=1}^s$. Then, $y_s(\mathbf{x})$ denotes the output of the most accurate and expensive model, while $y_1(\mathbf{x})$ is the output of the cheapest least accurate surrogate at our disposal. The autoregressive co-kriging scheme of [Kennedy and O'Hagan \(2000\)](#) reads as

$$Z_t(\mathbf{x}) = \rho_{t-1}(\mathbf{x})Z_{t-1}(\mathbf{x}) + \delta_t(\mathbf{x}), \quad t = 2, \dots, s, \quad (3.55)$$

where $\delta_t(\mathbf{x})$ is a Gaussian field independent of $\{Z_{t-1}(\mathbf{x}), \dots, Z_1(\mathbf{x})\}$, and distributed with expectation μ_{δ_t} and covariance Σ_t , i.e. $\delta_t \sim \mathcal{N}(\mu_{\delta_t}, \Sigma_t)$. Also, $\rho(\mathbf{x})$ is a scaling factor that quantifies the correlation between the model outputs $(y_t(\mathbf{x}), y_{t-1}(\mathbf{x}))$. In a Bayesian setting, $\rho(\mathbf{x})$ is treated as a random field with an assigned prior distribution that is later fitted to the data through inference. Here, to simplify the presentation we have assume that ρ is a deterministic scalar, independent of $\mathbf{x}$, and learned from the data through maximum likelihood estimation (see Sec. [3.6.2](#)).

The derivation of this model is based on the Markov property

$$\text{cov} \{Z_t(\mathbf{x}), Z_{t-1}(\mathbf{x}') | Z_{t-1}(\mathbf{x})\} = 0, \quad \forall \mathbf{x} \neq \mathbf{x}', \quad (3.56)$$

which, according to [Kennedy and O'Hagan \(2000\)](#), translates into assuming that given $Z_{t-1}(\mathbf{x})$, we can learn nothing more about $Z_t(\mathbf{x})$ from any other model output $Z_{t-1}(\mathbf{x}')$, for $\mathbf{x} \neq \mathbf{x}'$.

The resulting posterior distribution at the t^{th} co-kriging level has a mean and covariance given by

$$\begin{aligned} \mu_{Z_t|y_t, y_{t-1}, \dots, y_1}(\mathbf{x}) &= \mu_t + (\Sigma_t A_t^T)(A_t \Sigma_t A_t^T + \sigma_t I)^{-1}(y_t - A_t \mu_t), \\ \Sigma_{Z_t|y_t, y_{t-1}, \dots, y_1}(\mathbf{x}) &= \Sigma_{Z_t} + \rho_{t-1}^2 \Sigma_{Z_{t-1}} + \rho_{t-1}^2 \rho_{t-2}^2 \Sigma_{Z_{t-2}} + \dots + (\prod_{i=1}^{s-1} \rho_i)^2 \Sigma_{Z_1}, \end{aligned} \quad (3.57)$$

where μ_t is the mean value of $Z_t(\mathbf{x})$, Σ_t is a covariance matrix comprised by t blocks, representing all cross-correlations between levels $\{t, t-1, \dots, 1\}$, and $\Sigma_{Z_t} = \Sigma_t - (\Sigma_t A_t^T)(A_t \Sigma_t A_t^T + \sigma_t I)^{-1}(A_t^T \Sigma_t)$ is the covariance of the co-kriging predictor at level t . Also, σ_t is the variance of the noise at level t , and A_t is a simple matrix that restricts a realization of the Gaussian field $Z_t(\mathbf{x})$ to the locations of observed data $y_t(\mathbf{x})$ at level t . Similarly to ordinary kriging, the set of unknown parameters $\{\mu_t, \rho_{t-1}, \dots, \rho_1, \theta_t, \dots, \theta_1, \sigma_t, \dots, \sigma_1\}$ are determined from the data using machine learning techniques. To simplify the presentation, we have assumed that the scaling factor ρ attains scalar values, hence it does not depend on $\mathbf{x}$.

Although this co-kriging framework provides an elegant way of blending variable fidelity model output, it may easily become computationally infeasible if many levels of fidelity and/or a large number of data observations are available. The computational barrier is imposed by the need to repeatedly invert the full covariance matrix Σ_t at the model fitting stage, with the number of different fidelity levels resulting to

an increasing matrix size, while large observation data-sets introduce ill-conditioning. These are well known limiting factors for kriging but they become even more evident for co-kriging models where the total number of observations is the sum of the observations at all fidelity levels. However, recent findings of [Le Gratiet and Garnier \(2014\)](#) provide an equivalent formulation that overcomes the computational complexity issues of the Kennedy and O'Hagan model by decoupling the aforementioned s -level co-kriging scheme into s independent kriging problems. The resulting recursive scheme is proved to provide a predictive mean and variance that is identical to the coupled Kennedy and O'Hagan model, although it could potentially lead to a drastic reduction of the size and condition number of the covariance matrices that need to be inverted.

The key idea behind the derivation of [Le Gratiet and Garnier \(2014\)](#) is to replace the Gaussian process $Z_{t-1}(\mathbf{x})$ in Eq. 3.55, with a process $\tilde{Z}_{t-1}(\mathbf{x})$ that is conditioned by all the known observations $\{y_{t-1}, y_{t-2}, \dots, y_1\}$ up to level $(t-1)$, while assuming that the corresponding experimental design sets $(D_i)_{i=1}^{t-1}$ have a nested structure, i.e $D_1 \subseteq D_2 \subseteq \dots \subseteq D_{t-1}$. Then, the problem at the t^{th} -level reduces to an ordinary kriging predictor with mean and covariance given by

$$\begin{aligned} \mu_{Z_t|y_t, y_{t-1}, \dots, y_1}(\mathbf{x}) &= \rho_{t-1} \mu_{Z_{t-1}|y_{t-1}, \dots, y_1}(\mathbf{x}) + \mu_t + \\ &\quad + (\Sigma_t A_t^T)(A_t \Sigma_t A_t^T + \sigma_t I)^{-1} [y_t - A_t(\rho_{t-1} \mu_{Z_{t-1}|y_{t-1}, \dots, y_1}(\mathbf{x}) + \mu_t)], \\ \Sigma_{Z_t|y_t, y_{t-1}, \dots, y_1}(\mathbf{x}) &= \rho_{t-1}^2 \Sigma_{Z_{t-1}} + \Sigma_t - (\Sigma_t A_t^T)(A_t \Sigma_t A_t^T + \sigma_t I)^{-1} (A_t^T \Sigma_t). \end{aligned} \tag{3.58}$$

To underline the potential benefits of this approach, we note that the matrix Σ_t in the Kennedy and O'Hagan model (see Eq. 3.57) has size $\sum_{t=1}^s n_t \times \sum_{t=1}^s n_t$, where n_t is the number of observations at the t^{th} fidelity level. On the other hand, the recursive co-kriging approach involves the inversion of s covariance ma-

trices $(\Sigma_t)_{t=1}^s$ (see Eq. 3.58) of size $n_t \times n_t$, where n_t is the number of observations $y_t(\mathbf{x})$ at level t (see Le Gratiet and Garnier (2014)). Moreover, we note that at each recursive level, the number of unknown parameters to be learned from the data reduces to $\{\mu_t, \rho_{t-1}, \theta_t, \sigma_t\}$, compared to the large parametric set of $\{\mu_t, \rho_{t-1}, \dots, \rho_1, \theta_t, \dots, \theta_1, \sigma_t, \dots, \sigma_1\}$ of the coupled Kennedy and O'Hagan scheme.

3.6.2 Workflow

Algorithms 1-4 summarize the main implementation aspects of the proposed multi-level co-kriging framework, formulated using sparse GMRF precision matrices. Namely, we provide pseudocode for the main co-kriging recursion loop (see Algorithm 1), the hyperparameter learning through maximum likelihood estimation (see Algorithm 2), the computation of the negative log-likelihood function of the co-kriging posterior (see Algorithm 3), and the computation of the co-kriging variance (see Algorithm 4).

Algorithm 1 Multi-level recursive co-kriging with GMRFs

```

% Loop through all recursion levels (lowest to highest fidelity)
for  $t = 1 : s$  do
    % Load input design variables and observations at level  $t$  ( $\mathbf{x} \in D_t$ )
     $\mathbf{x} \leftarrow$  input design data and prediction points
     $A_t \leftarrow$  simple restriction matrix (see Sec:3.6.1.1)
     $y_t(A_t \mathbf{x}) \leftarrow$  observed data
    % Compute the mass and stiffness matrices of a finite element (FEM)
    % discretization of the design space
     $[M, S] \leftarrow$  FEM_discretization()
    % Learn the co-kriging hyperparameters  $\{\mu_t, \rho_{t-1}, \kappa_t, \tau_t, \sigma_t\}$  through MLE
     $\{\theta^{\text{opt}}, \mu^{\text{opt}}, Q_c^{-1}\} \leftarrow$  MLE_optimization( $\mathbf{x}, y_t, A_t, M, S, t$ )
    % Evaluate the co-kriging predictor and variance at level  $t$  from Eq.3.58
    % and proceed to the next level  $t = t + 1$ 
     $\mu_{Z_t|y_t, y_{t-1}, \dots, y_1} \leftarrow \mu^{\text{opt}}$ 
     $\Sigma_{Z_t|y_t, y_{t-1}, \dots, y_1} \leftarrow$  Variance( $\theta^{\text{opt}}, Q_c^{-1}, t$ )
% Finally, return the co-kriging predictor and variance at the highest level of fidelity
return  $\mu_{Z_s|y_s, y_{s-1}, \dots, y_1}, \Sigma_{Z_s|y_s, y_{s-1}, \dots, y_1}$ 

```

Algorithm 2 Maximum likelihood estimation for hyperparameter learning

```

procedure MLE_OPTIMIZATION( $\mathbf{x}, y_t, A_t, M, S, t$ )
  if  $t > 1$  then
     $y^* \leftarrow y_t - \rho_{t-1} A_t \mu_{Z_{t-1}|y_{t-1}, y_{t-2}, \dots, y_1}$ 
     $\mu_t \leftarrow \mu_t + \rho_{t-1} \mu_{Z_{t-1}|y_{t-1}, y_{t-2}, \dots, y_1}$ 
     $\theta^* \leftarrow \{\mu_t, \rho_{t-1}, \kappa_t, \tau_t, \sigma_t\}$ 
  else
     $y^* \leftarrow y_t$ 
     $\theta^* \leftarrow \{\mu_t, \kappa_t, \tau_t, \sigma_t\}$ 
  % Learn the optimal hyperparameters that minimize the negative
  % log-likelihood of the co-kriging posterior
   $\{\theta^{\text{opt}}, \mu^{\text{opt}}, Q_c^{-1}\} \leftarrow \min_{\theta^*} \text{NegLogLikelihood}(\theta^*, y^*, A_t, M, S)$ 
  return  $\{\theta^{\text{opt}}, \mu^{\text{opt}}, Q_c^{-1}\}$ 

```

Algorithm 3 Computation of the negative log-likelihood of the co-kriging posterior

```

procedure NEGLOGLIKELIHOOD( $\theta^*, y^*, A_t, M, S$ )
  % Construct the discrete Helmholtz operator of the SPDE (see Eq. 3.12)
   $H \leftarrow \frac{1}{\tau_2} (\kappa_t^2 M + S)$ 
  % Construct the sparse precision matrix  $Q_t$  (see Eq. 3.21),
  % and the diagonal noise precision  $Q_{\mathcal{E}}$ 
   $Q_t \leftarrow H M^{-1} H$ 
   $Q_{\mathcal{E}} \leftarrow \frac{1}{\sigma_t} I$ 
  % Factorize  $Q_t = U U^T$  using the Cholesky decomposition
   $U \leftarrow \text{Cholesky}(Q_t)$ 
  % Compute the conditional precision  $Q_c$ 
   $Q_c \leftarrow Q_t + A_t^T Q_{\mathcal{E}} A_t$ 
  % Compute the conditional predictor  $\mu^*$ , using the Woodbury matrix identity
  % and Cholesky back-substitution to efficiently compute  $Q_c^{-1}$ 
   $\mu^* \leftarrow \mu_t + Q_c^{-1} A_t Q_{\mathcal{E}} (y^* - A_t \mu_t)$ 
   $\text{NegLogLikelihood} \leftarrow -\frac{1}{2} \{ \log |Q_t| + \log |Q_{\mathcal{E}}| - \log |Q_c| -$ 
     $- [(\mu^* - \mu_t)^T Q_t (\mu^* - \mu_t) + (y^* - A_t \mu^*)^T Q_{\mathcal{E}} (y^* - A_t \mu^*)] \}$ 
  Return  $\{\text{NegLogLikelihood}, \mu^*, Q_c^{-1}\}$ 

```

Algorithm 4 Computation of the co-kriging variance

```

procedure VARIANCE( $\theta^{\text{opt}}, Q_c^{-1}, t$ )
  if  $t > 1$  then
    return  $\text{diag}(\rho_{t-1}^2 \Sigma_{Z_{t-1}|y_{t-1}, y_{t-2}, \dots, y_1} + Q_c^{-1})$ 
  else
    return  $\text{diag}(Q_c^{-1})$ 

```

3.7 Results

In this section we present representative results for using the proposed multi-fidelity information fusion with GMRFs in different contexts. First, we test the accuracy of the nodal spectral element basis, and consider a problem of gappy field reconstruction from a deterministic 2D Helmholtz equation. Next, we present results on a prototype benchmark in risk-averse design, where a stochastic optimization problem is solved to identify the optimal design of a structural element. Then, we consider a test problem involving the regression of the mean field of a random function when a family of low-fidelity representations is available. Finally, we conclude with an uncertainty quantification problem in fluid mechanics that involves stochastic laminar wakes past a circular cylinder.

3.7.1 Helmholtz equation in 2D

3.7.1.1 Convergence rate

First, we demonstrate the exponential convergence rate of the nodal spectral element discretization described in Sec. 3.4. To this end, we consider the Helmholtz equation in a bounded two-dimensional domain Ω

$$\nabla^2 u - \lambda u = f, \tag{3.59}$$

where λ is a prescribed constant, f is a known forcing term, and u is the unknown scalar field we wish to solve for. At the domain boundary, we assume Dirichlet boundary conditions $u|_{\partial\Omega} = u_0(\mathbf{x})$. Here, we have considered a square domain, $\Omega =$

$\{(x, y) : 0 \leq x \leq 2, -1 \leq y \leq 1\}$, discretized using four equi-spaced quadrilateral spectral elements, with 2 elements in each direction. Choosing $\lambda = 100$ and $f = -(\lambda + 2\pi^2) \cos(\pi x) \cos(\pi y)$ we can obtain the analytic solution to the Helmholtz problem as

$$u = \cos(\pi x) \cos(\pi y) \quad (3.60)$$

Using the same polynomial expansion order in each element, we vary the polynomial order systematically and compute the corresponding L^∞ and L^2 errors of the numerical solution against the analytical solution. At the domain boundaries the numerical solution is set to the analytic expression of Eq. 3.60. Figure 3.2 shows the computed numerical solution for polynomial expansion order $P = 5$, and the corresponding exponential error decay as a function of the element order.

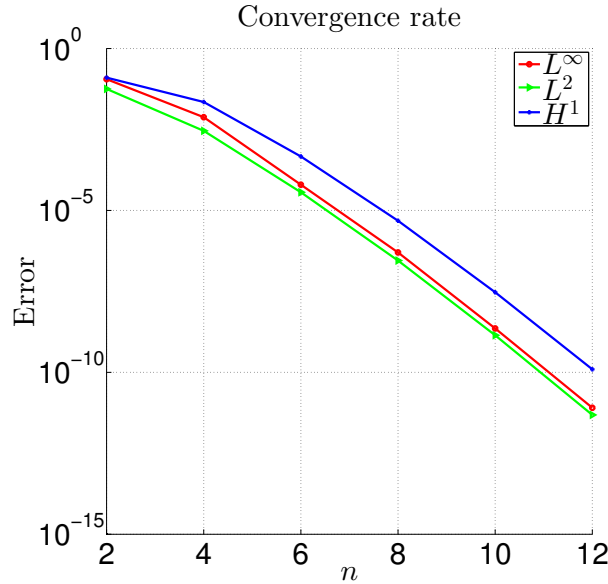


Figure 3.2: *Helmholtz equation in 2D*: Exponential convergence rate of the nodal spectral element discretization for a Helmholtz equation in 2D. The error between the numerical and the exact solution has been computed for different numbers of expansion modes n , using the L^∞ , L^2 , and H^1 norms, respectively.

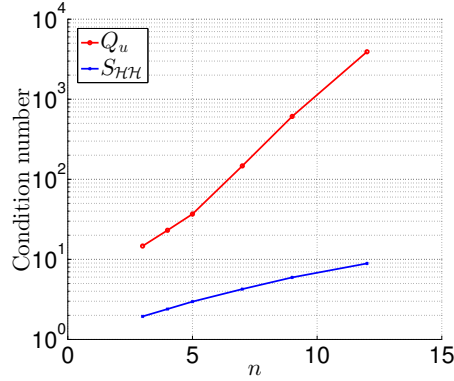
3.7.1.2 Gappy field reconstruction using GMRF kriging

Having demonstrated the convergence properties of the nodal spectral element basis, we now focus our attention on employing the observation model of Eq. 3.1 for performing reconstruction of gappy fields. In what follows, we consider the case of *ordinary* kriging, in which the mean of the latent model $u(\mathbf{x})$ is assumed to be unknown but independent of $\mathbf{x}$, although any other form of kriging is also directly applicable. The latent process $u(\mathbf{x})$ is assumed to satisfy the SPDE of Eq 3.6, possibly subject to any *a-priori* known boundary conditions. As discussed in Sec. 3.4, the SPDE under the nodal spectral element discretization generates a GMRF with sparse precision.

Again, we consider the 2D Helmholtz problem of Eq. 3.59, for which an exact solution is readily available. The goal here is to train the prediction model of Eq. 3.1 based on a set of known observations such that it is able to generate a reasonably accurate reconstruction of the solution to the Helmholtz problem. Here, we introduce gappyness in the interior of the domain, while the SPDE satisfies the exact solution through Dirichlet boundary conditions at all four boundaries of the square. Parameter estimation has been performed using the Maximum Likelihood Estimation (MLE) methodology outlined in Sec. 3.5.1.

Figures 3.3-3.6 and Table 3.1 show the size, structure, sparsity pattern and condition number of the computed mass, stiffness, precision and homogeneous-homogeneous Schur complement matrices for different number of expansion modes n . The results correspond to 4-element quadrilateral nodal spectral element discretization of the SPDE (Eq. 3.6). We observe that the size and condition number of $S_{\mathcal{H}\mathcal{H}}$ exhibit good scaling with increasing polynomial order and are considerably smaller than the condition number and size of the precision Q_u . This fact may favor the block co-

variance approach suggested in Sec. 3.4 for large problems where the large condition number of Q_u may prohibit the use of Cholesky decomposition, as the covariance can then be computed by inverting the homogeneous-homogeneous Schur complement $S_{\mathcal{H}\mathcal{H}}$.



n	Q_u	$S_{\mathcal{H}\mathcal{H}}$
3	25	5
4	49	9
5	81	13
7	169	21
9	289	29
12	529	41

Figure 3.3: *Helmholtz equation in 2D*: Condition number of the precision Q_u and homogeneous-homogeneous Schur complement $S_{\mathcal{H}\mathcal{H}}$ for different number of expansion modes n .

Table 3.1: *Helmholtz equation in 2D*: Leading size dimension of the precision Q_u and homogeneous-homogeneous Schur complement $S_{\mathcal{H}\mathcal{H}}$ matrices for different number of expansion modes n .

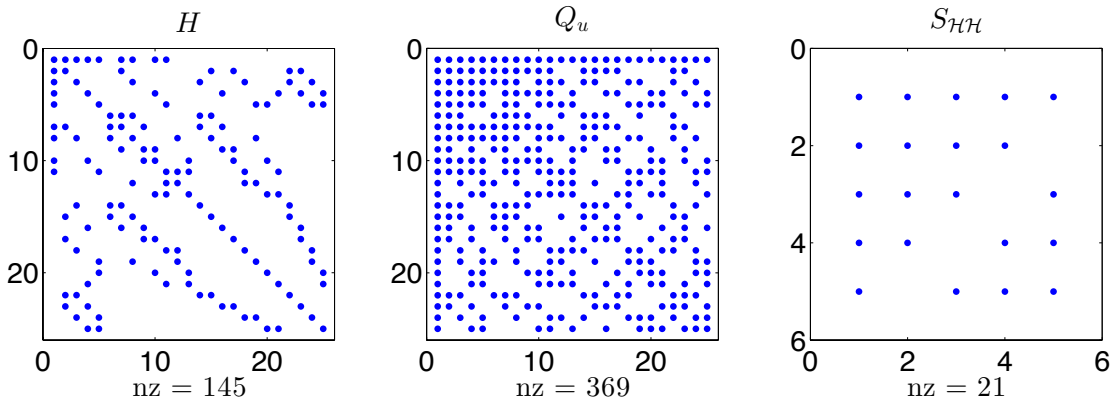


Figure 3.4: *Helmholtz equation in 2D*: Sparsity patterns of the Helmholtz, precision and homogeneous-homogeneous Schur complement matrices for a second-order polynomial expansion basis ($n = 3$).

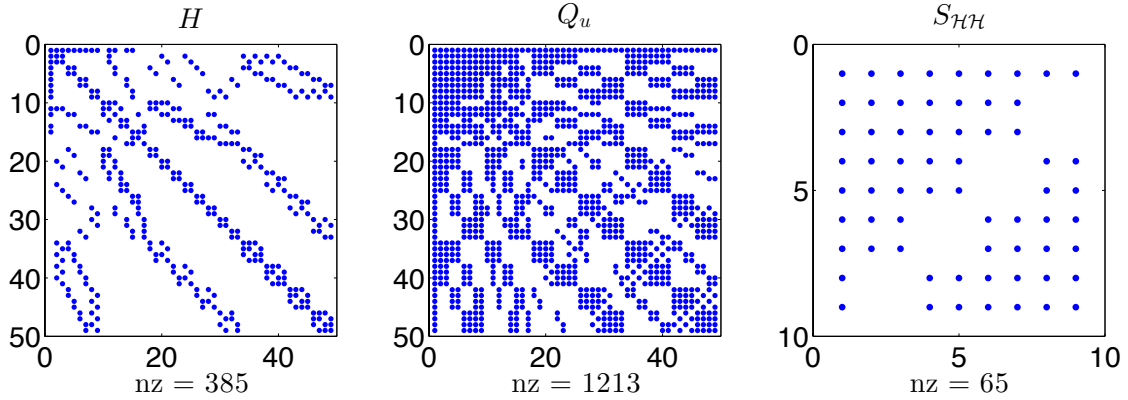


Figure 3.5: *Helmholtz equation in 2D*: Sparsity patterns of the Helmholtz, precision and homogeneous-homogeneous Schur complement matrices for a third-order polynomial expansion basis ($n=4$).

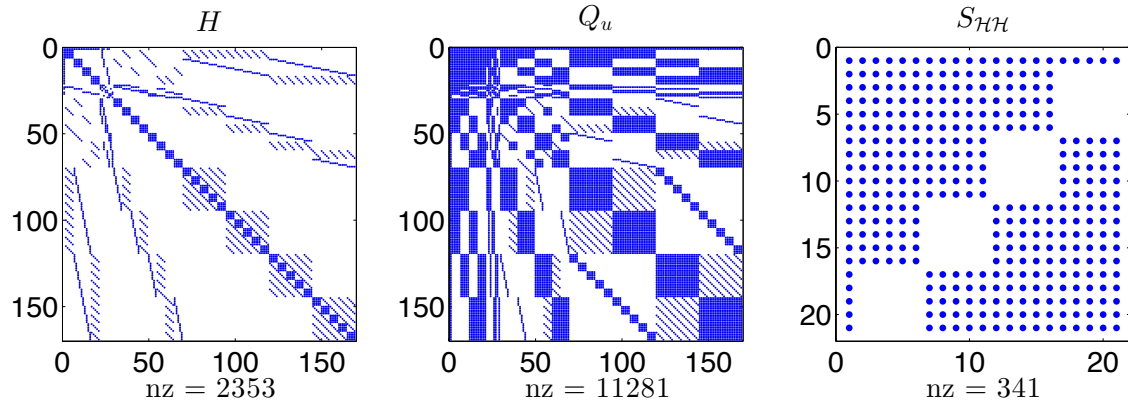


Figure 3.6: *Helmholtz equation in 2D*: Sparsity patterns of the Helmholtz, precision and homogeneous-homogeneous Schur complement matrices for a sixth-order polynomial expansion basis ($n=7$).

In Fig. 3.7 we present the effect of increasing the the number of modes in the basis expansion on the accuracy of the resulting kriging predictor. To this end we fix a single observation point located in $(0,1)$ and increase the polynomial order of the nodal basis expansion that discretizes Eq. 3.6. We observe that as the polynomial order is gradually increased from 3 to 9, the kriging predictor becomes up to 3 orders of magnitude more accurate, clearly indicating the benefits of high-order approximation. Moreover, note that as the generated predictor automatically satis-

fies the boundary conditions, only one observation point at the center of the domain is sufficient to accurately reconstruct the unknown field. Alternatively, in Fig. 3.7 we fix the number of expansion modes to $n = 7$, and we observe the increase of the predictor's accuracy, accompanied by the decrease in the kriging variance, as we add randomly sampled observations.

3.7.2 A prototype problem in risk-averse design

Consider a rectangular cross section of a short structural column, with depth x_1 and width x_2 , under uncertain yield stress and uncertain bending and axial loads. Assuming an elastic-perfectly plastic material, a limit state function that quantifies a relationship between loads and capacity is described by the random field (see [Kuschel and Rackwitz \(1997\)](#))

$$Y(x_1, x_2; \xi_1, \xi_2, \xi_3) = \frac{4\xi_1}{x_1^2 x_2 \xi_3} + \frac{\xi_2^2}{x_1^2 x_2^2 \xi_3^2} - 1, \quad (3.61)$$

where the bending moment load ξ_1 , the axial load ξ_2 , and the material yield stress ξ_3 are distributed as

$$\xi_1 \sim \mathcal{N}(2000, 400)$$

$$\xi_2 \sim \mathcal{N}(500, 100)$$

$$\xi_3 \sim \log \mathcal{N}(5, 0.5).$$

Note that lower values of Y indicate more favorable capacity-to-load ratios, while high values correspond to critical regimes where failure may occur. Our aim here is to identify an optimal configuration (x_1^*, x_2^*) in the two-dimensional design space

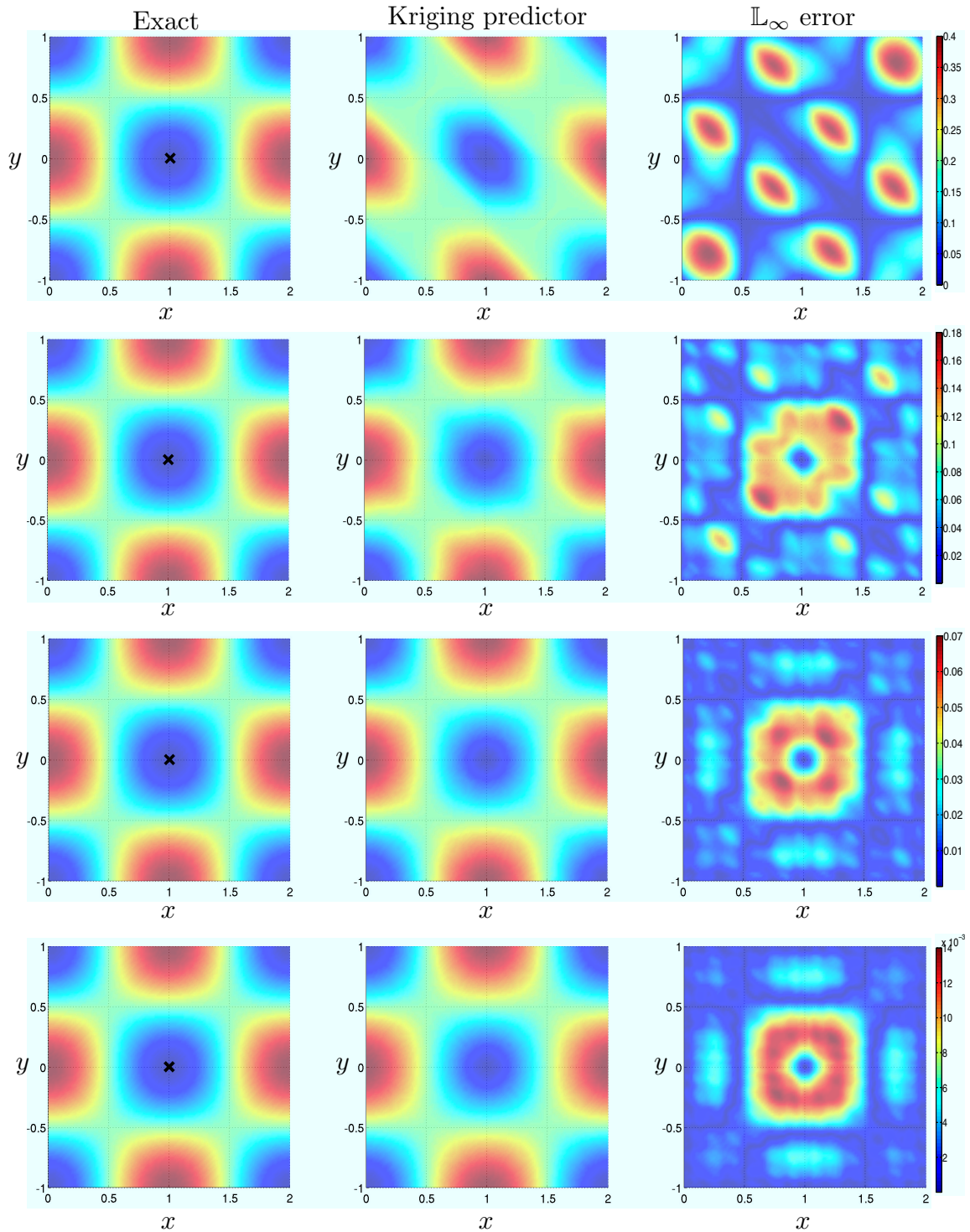


Figure 3.7: *Helmholtz equation in 2D*: Observation points, kriging prediction and pointwise L^∞ error for a single observation point located at $(1,0)$. *Top to bottom*: Nodal spectral element discretization with 4 quadrilateral elements and 3,5,7, and 9 expansion modes, respectively.

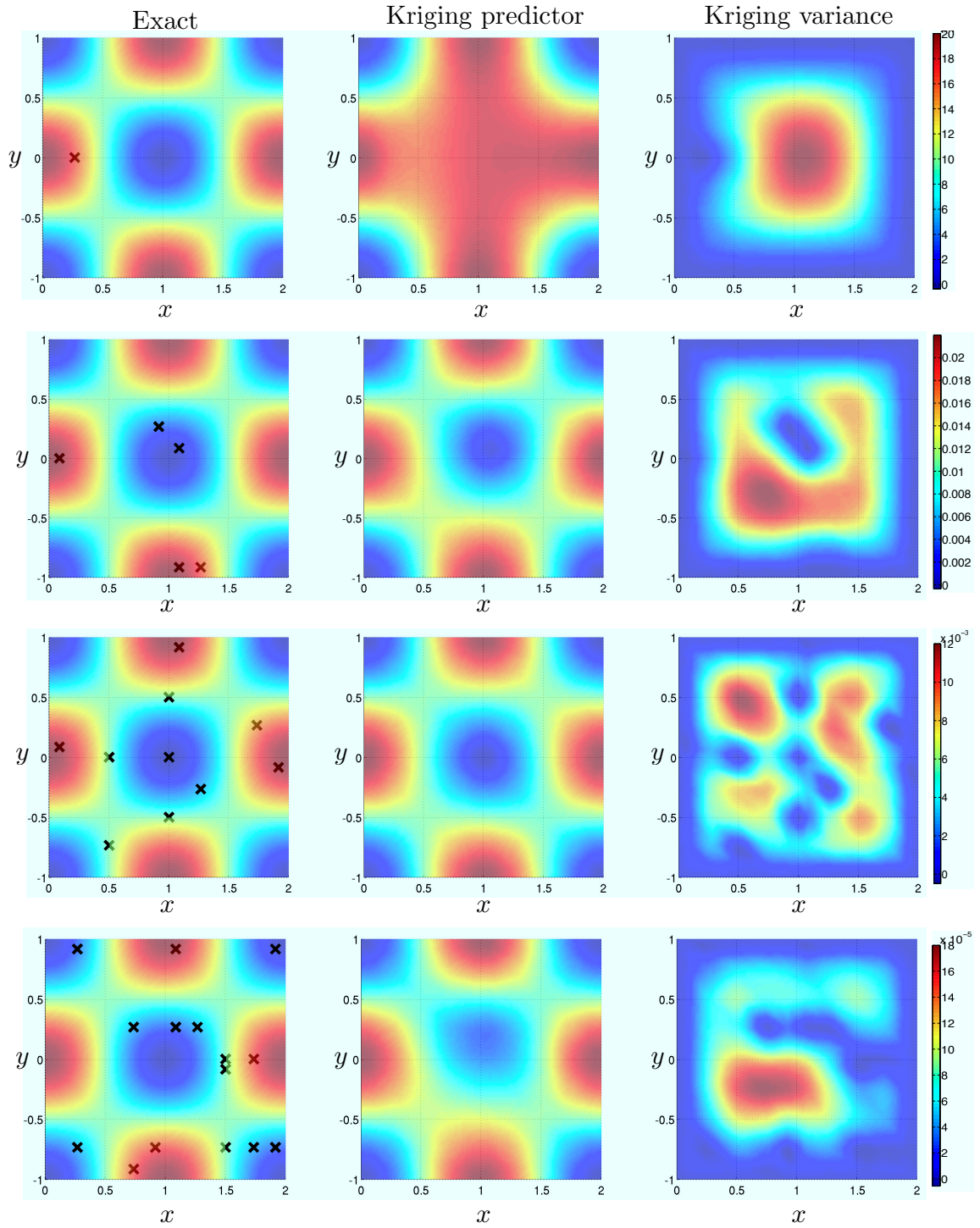


Figure 3.8: *Helmholtz equation in 2D*: Observation points, kriging prediction and kriging variance for a nodal spectral element discretization with 4 quadrilateral elements and $n = 7$ expansion modes. *Top to bottom*: Increasing the number of randomly selected observations from 1,5,10, to 15, respectively.

$D_1 \times D_2 = [2, 20] \times [2, 10]$ that satisfies a set of prescribed criteria. Here, we let this goal translate to identifying a design that minimizes the area of the column with the associated design *risk* staying below a given threshold c . To quantify the risk associated with a given response $Y(\mathbf{x}; \xi)$, we employ a *superquantile risk measure*, which for a parameter $\alpha \in [0, 1]$ is defined as (see [Rockafellar and Royset \(2014\)](#)):

$$\mathcal{R}_\alpha(Y(\mathbf{x}; \xi)) = \min_{c \in \mathbb{R}} \left\{ c + \frac{1}{1 - \alpha} \mathbb{E}[\max\{Y(\mathbf{x}; \xi) - c, 0\}] \right\}. \quad (3.62)$$

For $\alpha = 0$, the superquantile risk measure of a continuously distributed random field $Y(\mathbf{x}; \xi)$, reduces to the expectation operator, i.e., $\mathcal{R}_0 = \mathbb{E}[Y(\mathbf{x}; \xi)]$, thus characterizing a risk-neutral situation. However, for non-zero values $\alpha \in (0, 1)$, the superquantile risk of $Y(\mathbf{x}; \xi)$ returns the average of the $(1 - \alpha)\%$ highest realizations of $Y(\mathbf{x}; \xi)$, therefore informing a *risk-averse* decision making mindset (see [Rockafellar and Royset \(2014\)](#)).

The design problem can be formulated in a constrained minimization framework which reads as

$$\begin{aligned} & \min_{x_1 \in D_1, x_2 \in D_2} x_1 x_2 \\ & \text{subject to } \mathcal{R}_\alpha(Y) \leq c \end{aligned} \quad (3.63)$$

The expectation operator $\mathbb{E}[\max\{Y(\mathbf{x}; \xi) - c, 0\}]$ appearing in the computation of the superquantile risk $\mathcal{R}_\alpha(Y(\mathbf{x}; \xi))$ can be approximated using a variety of different probabilistic methods of different cost and fidelity. Here, we choose to discretize the three dimensional parametric space of (ξ_1, ξ_2, ξ_3) using two different models. The high-fidelity model is a probabilistic collocation method (PCM) (see [Foo et al. \(2008\)](#)) on a tensor product grid with 10 quadrature points along each random dimension (total of 10^3 quadrature points), while the low-fidelity model is a Smolyak sparse grid level-2 quadrature (SG-L2) (see [Novak and Ritter \(1996\)](#)) with a total of only

58 quadrature points.

This setup yields a simple application of the proposed multi-fidelity methodology, where we have a single model in physical space (Eq. 3.61) and two levels of fidelity in probability space (PCM, SG-L2). Employing a two-level recursive co-kriging algorithm (see Sec. 3.6.1.1) we can explore the correlations between the two probabilistic models, and build an accurate representation of the target response surface $\mathcal{R}_\alpha(Y)$ by mainly utilizing the realizations of the lower-fidelity model (SG-L2, 120 points in design space), guided by few realizations of the higher fidelity model (PCM, 10 points in design space).

In Fig. 3.9, we present the computed response surface $\mathcal{R}_{0.8}(Y)$ as a function of the design variables (x_1, x_2) considering the 20% highest realizations of Y . The exact solution corresponds to computing $\mathcal{R}_{0.8}(Y)$ using 10^6 Monte Carlo samples (see Binder and Heermann (2010)). The resulting response surface becomes singular as $(x_1 \rightarrow 0, x_2 \rightarrow 0)$, while it attains a very flat profile in the rest of the design space. This flat behavior suggests that the identification of an optimal design as a solution to Eq. 3.63 is highly sensitive to the accuracy of the computed response surface. Here, we have chosen a design criterion of $\mathcal{R}_{0.8}(Y) \leq -0.99$, yielding an optimal solution for $(x_1^* = 13.94, x_2^* = 5.07)$ that does not depend on further resolution refinement.

Although this is a pedagogical prototype case in risk-averse design under uncertainty, it does not fully reflect the strengths of the proposed multi-fidelity framework as the two probabilistic methods (PCM, SG-L2) both happen to produce very similar estimations of the mean $\mathbb{E}[Y]$. The next example illustrates better the capabilities of the proposed methodology in the context of multi-fidelity regression of a random function.

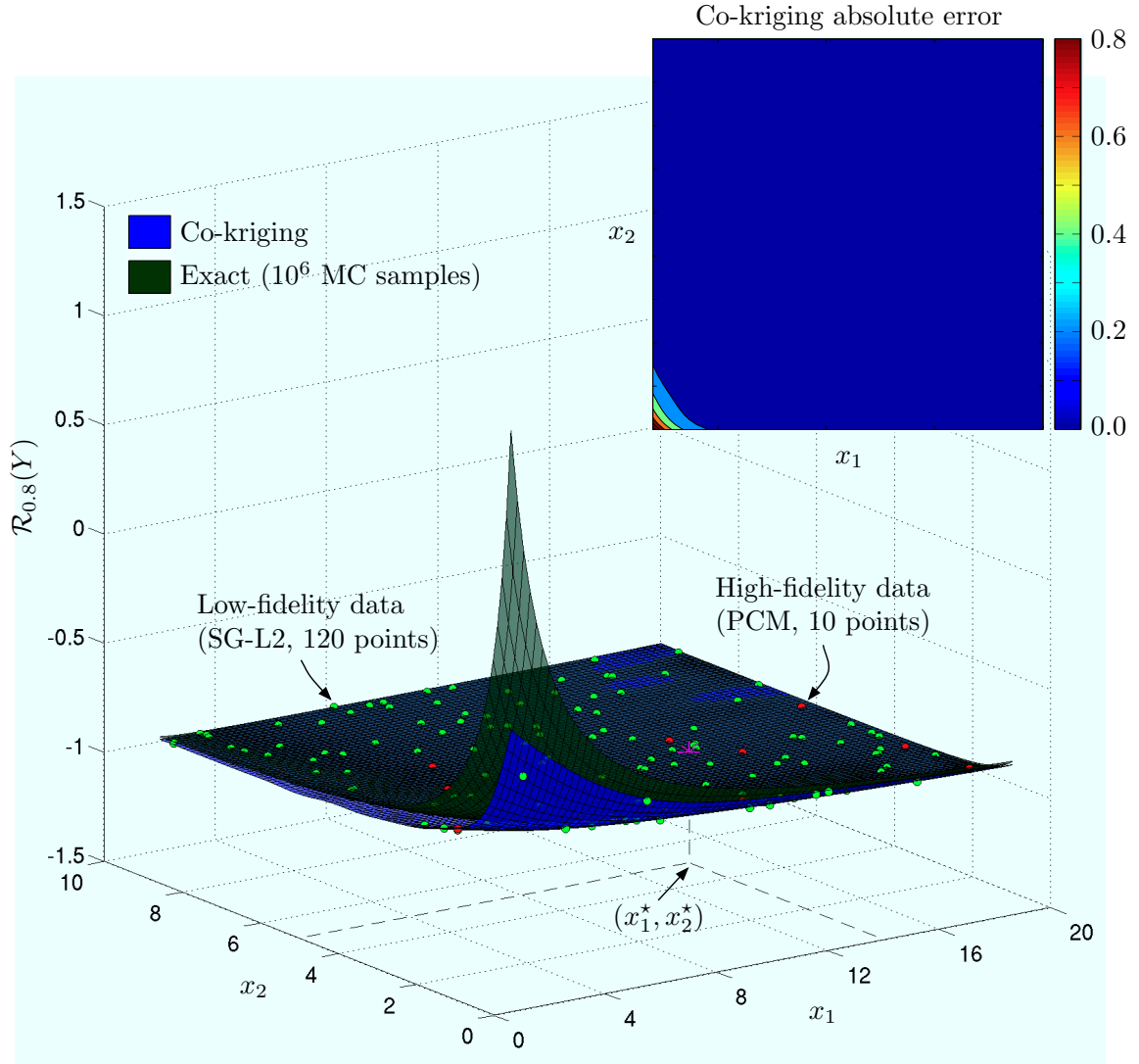


Figure 3.9: *A prototype problem in risk-averse design:* Exact response surface $\mathcal{R}_{0.8}(Y)$ and co-kriging predictor constructed using 120 low-fidelity (SG-L2) and 10 high-fidelity (PCM) observations. (x_1^*, x_2^*) denotes the optimal solution to the optimization problem of Eq. 3.63, while the inset plot shows the point-wise absolute error of the co-kriging predictor.

3.7.3 Multi-fidelity regression of a random function

Here, we consider a system with two input design variables $\mathbf{x} = (x_1, x_2)$, subject to external uncertainties, described by four standard normal random variables $\xi = (\xi_1, \xi_2, \xi_3, \xi_4)$. Let the response of this system, denoted by $Y(\mathbf{x}; \xi)$ be described by the random function

$$f_e(\mathbf{x}; \xi) = \xi_1 \sin^2(5\xi_1 x_1 + 2\xi_2 \xi_3 x_2) + 2\xi_4^2 e^{-x_1(\xi_2 + \xi_3)(x_2 - 0.5)^2} \cos^2(4x_1 + x_2). \quad (3.64)$$

Now, assume that $f_e(\mathbf{x}; \xi)$ returns the real high-fidelity response but is expensive to evaluate. On the other hand, let $f_c(\mathbf{x}; \xi)$ be a low-fidelity cheap to evaluate surrogate model that we can sample extensively as

$$f_c(\mathbf{x}; \xi) = 1.7f_e(\mathbf{x}; \xi) + 2\xi_1 \xi_2 \sin(x_1 + x_2) + 5\xi_3 \xi_4^2 e^{-x_1} \sin(x_1 + 7x_2). \quad (3.65)$$

Our goal here is to employ the proposed multi-fidelity framework to construct the response surface of the mean field $S(\mathbf{x}) = \mathbb{E}[Y(\mathbf{x}; \xi)]$. In order to approximate the expectation operator, we employ two methods of different fidelity in probability space. To this end, we choose our high-fidelity probabilistic method to be a Smolyak sparse grid level-5 quadrature (SG-L5) (see [Novak and Ritter \(1996\)](#)) that discretizes the four-dimensional parameter space using 4,994 quadrature points. Similarly, the low-fidelity method in probability space is a coarser, level-2 sparse grid quadrature (SG-L2) with just 57 quadrature points (see [Novak and Ritter \(1996\)](#)).

Therefore, our multi-fidelity setup consists of two models in physical space ($f_e(\mathbf{x}; \xi)$, $f_c(\mathbf{x}; \xi)$), and two models in probability space (SG-L5, SG-L2). This results in a family of re-

sponse surfaces S_{ij} , that can be organized as

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} \mathbb{E}_{SG-L2}[f_c(\mathbf{x}; \xi)] & \mathbb{E}_{SG-L5}[f_c(\mathbf{x}; \xi)] \\ \mathbb{E}_{SG-L2}[f_e(\mathbf{x}; \xi)] & \mathbb{E}_{SG-L5}[f_e(\mathbf{x}; \xi)] \end{pmatrix} \quad (3.66)$$

Fig. 3.10 demonstrates the response surface produced using a 4-level recursive co-kriging scheme, traversing the available models and data in the order $S_{11} \rightarrow S_{12} \rightarrow S_{21} \rightarrow S_{22}$, i.e., from lowest to highest fidelity. The resulting response surface is compared with an “exact” solution that is obtained by Monte Carlo integration of the high-fidelity physical model $f_e(\mathbf{x}; \xi)$ using 10^6 samples. Evidently, the highly non-linear response of the mean field is captured remarkably well by just using five observations of the expensive highest-fidelity model S_{22} , supplemented by a number of inaccurate low-fidelity observations from (S_{11}, S_{12}, S_{21}) . This observation is further confirmed by the uncertainty of the predictor, quantified by the co-kriging variance (see Fig. 3.10, inset), which is bounded by 10^{-3} .

3.7.4 Stochastic incompressible flow

We consider two-dimensional (2D) unsteady incompressible flow past a circular cylinder of diameter D , subject to a random inflow boundary condition. The flow is governed by the Navier-Stokes equations subject to the incompressibility constraint

$$\begin{cases} \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot (\nabla \mathbf{v}) = -\nabla p + \nu \nabla^2 \mathbf{v} \\ \nabla \cdot \mathbf{v} = 0 \end{cases} \quad (3.67)$$

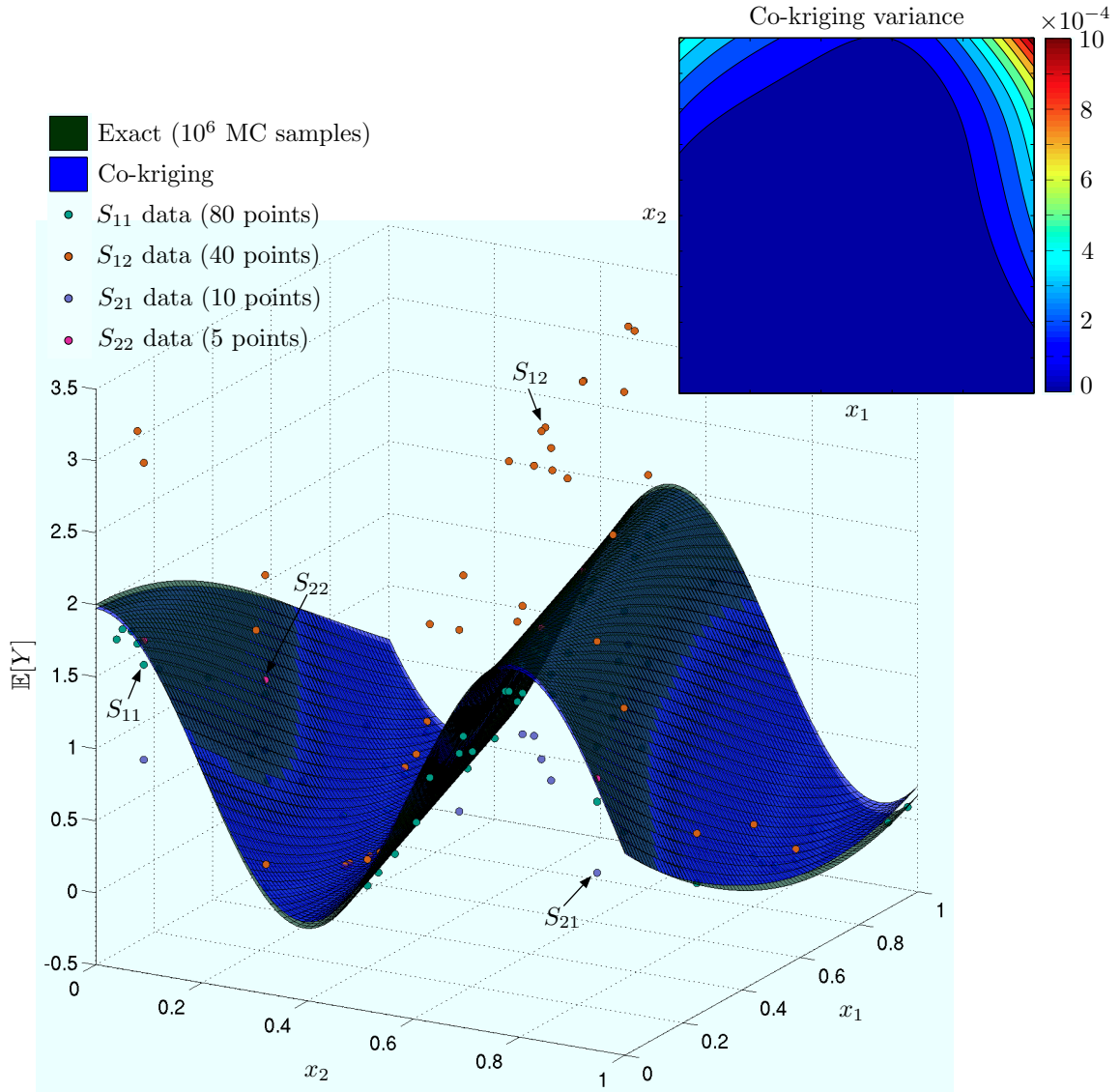


Figure 3.10: *Multi-fidelity regression of a random function*: Exact response surface $S = \mathbb{E}[Y(\mathbf{x}; \xi)]$ and co-kriging predictor constructed using four levels of fidelity: 80 S_{11} points ($f_c(\mathbf{x}; \xi)$, SG-L2), 40 S_{12} points ($f_c(\mathbf{x}; \xi)$, SG-L5), 10 S_{21} points ($f_e(\mathbf{x}; \xi)$, SG-L2), and 5 S_{22} points ($f_e(\mathbf{x}; \xi)$, SG-L5). The inset plot shows the point-wise variance of the co-kriging predictor.

where $\mathbf{v}$ is the fluid velocity vector, p is the fluid pressure, t is time, and ν is the kinematic viscosity of the fluid. The system is discretized in space using the spectral/ hp element method (SEM), according to which the computational domain is decomposed into a set of polymorphic non-overlapping elements (see [Karniadakis and Sherwin \(2013\)](#)). Within each element, the solution is approximated as a linear combination of hierarchical, mixed-order, semi-orthogonal Jacobi polynomial expansions [Karniadakis and Sherwin \(2013\)](#). Temporal integration is based on decoupling the velocity and pressure by applying a high-order time-splitting scheme (see [Karniadakis and Sherwin \(2013\)](#)).

The random inflow has a parametric expression of the form

$$U_{\infty}(\sigma_1, \sigma_2; \xi_1, \xi_2) = 1 + \sigma_1 \sin\left(\frac{\pi y}{9}\right) + \sigma_2 \left[\xi_1 \sin\left(\frac{\pi y}{9}\right) + \xi_2 \cos\left(\frac{\pi y}{9}\right) \right]. \quad (3.68)$$

where (σ_1, σ_2) are parameters controlling the amplitude of the skewness of the inflow noise (design variables), y is the coordinate transverse to the flow, and (ξ_1, ξ_2) are random variables with standard normal distributions.

Our goal here is to construct the response surface for the 0.6-superquantile risk of the base pressure coefficient $\mathcal{R}_{0.6}(C_{BP})$ (see Fig. [3.11](#)) at Reynolds number $Re = \frac{U_{\infty} D}{\nu} = 100$.

The proposed multi-fidelity modeling framework is illustrated here considering a single model in physical space and three models in probability space. The physical model returns realizations of the flow field solution produced by direct numerical simulations of Eq. [3.67](#) using the spectral/ hp element method. All simulations are started from a zero-velocity initial condition, and are integrated until a stable limit-cycle state is reached. This state is characterized by the well known von-Karman

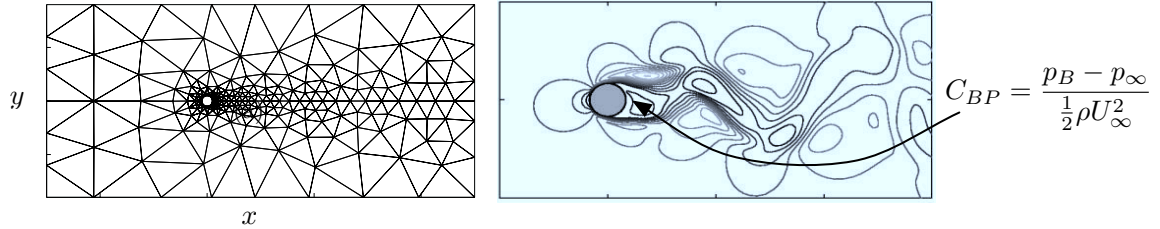


Figure 3.11: *Incompressible stochastic flow past a cylinder:* (Left) The computational mesh consists of 412 triangular spectral elements of polynomial order 11. (Right) Standard deviation of the stream-wise velocity component at time $t = 4$. The quantity of interest here is the base pressure coefficient C_{BP} , where p_B denotes the fluid pressure on the cylinder surface at 180° from the stagnation point (base pressure), while U_∞ and p_∞ are the free-stream velocity and pressure, respectively.

vortex shedding pattern in which the unsteady separation of the flow around the cylinder gives rise to a time-periodic base pressure signal. We choose three models in probability space to compute the expectation operator in Eq 3.62. The highest fidelity model is probabilistic collocation on a tensor product grid (PCM) (see Foo et al. (2008)), the intermediate fidelity is a Smolyak sparse grid level-2 quadrature (SG-L2) Novak and Ritter (1996), and lowest fidelity model is Monte Carlo integration (MC) (see Binder and Heermann (2010)). For each model, we construct a uniform grid in the (σ_1, σ_2) -plane, and perform an ensemble of simulations at each grid-point according to the corresponding sampling strategy for integrating out (ξ_1, ξ_2) (see Table 3.2). Fig. 3.12 illustrates the PCM samples of the inflow boundary condition for different values of (σ_1, σ_2) . This uncertainty quantification task results to a total of 5,526 Navier-Stokes simulations that were performed in parallel on one rack of IBM BG/Q (16,384 cores) with a CPU time of about 40 minutes.

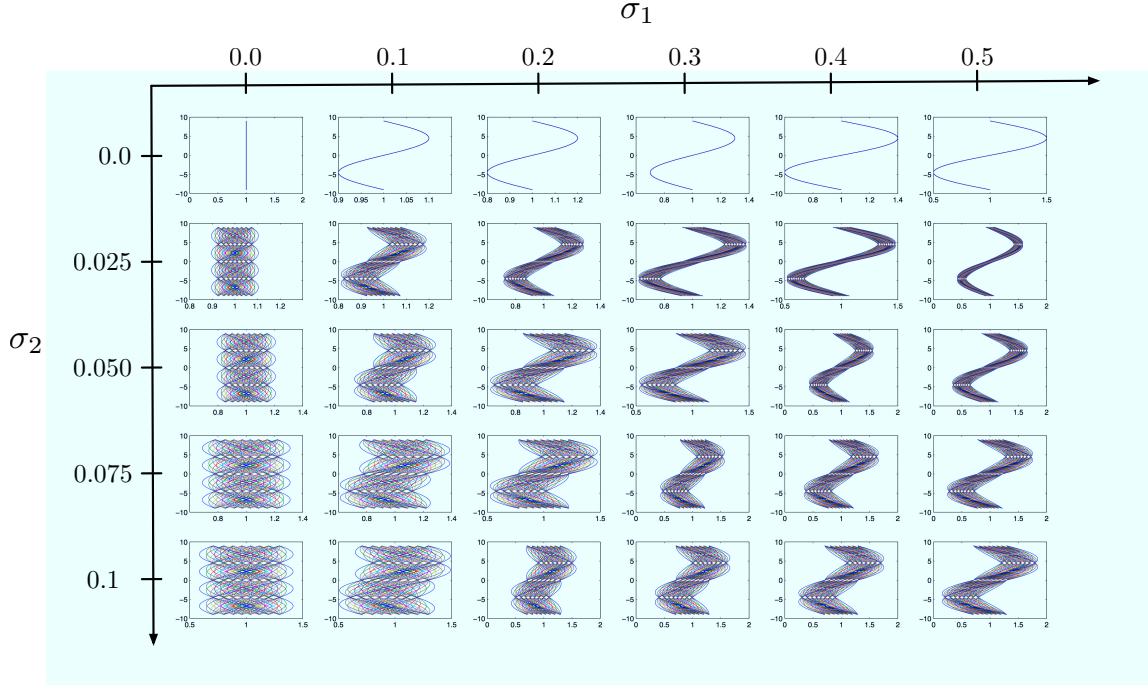


Figure 3.12: *Incompressible stochastic flow past a cylinder*: Probabilistic collocation (PCM) samples of the velocity inflow profile on a uniform 6×5 (σ_1, σ_2) -grid. At each grid point, a total of 64 samples is used to integrate out the randomness introduced by (ξ_1, ξ_2) in Eq. 3.68. Note that σ_1 controls the skewness of the inflow profile, while σ_2 acts as a noise amplifier.

Method	(σ_1, σ_2) -grid	Samples/grid-point	Total samples
PCM	6×5	64	1920
SG-L2	11×9	22	2178
MC	21×17	4	1428

Table 3.2: Stochastic incompressible flow: Size of the uniform discretization grid in the (σ_1, σ_2) -plane, and number of samples per grid-point for each of the probabilistic methods employed.

Fig. 3.13 shows the response surface for the 0.6-superquantile risk of the base pressure coefficient $\mathcal{R}_{0.6}(C_{BP})$, computed using a 3-level recursive co-kriging. The predictive scheme is trained using 100 low-fidelity (MC), 15 intermediate fidelity (SG-L2), and only 5 high-fidelity (PCM) observation points in the (σ_1, σ_2) -plane. The

accuracy of the co-kriging predictor is assessed against an “exact” response surface constructed by using 357 uniformly distributed observations of the highest fidelity model (PCM), returning an error of 0.051 in the $\mathbb{L}_2$ norm. More importantly, the uncertainty of the co-kriging predictor is negligible, as its variance is firmly bounded below 3×10^{-5} . Note that if we only used a 1-level kriging scheme trained on the 5 high-fidelity observations, the resulting $\mathbb{L}_2$ is 3.536. It is evident that incorporating information from lower fidelity sources can drastically improve the accuracy of the metamodel.

3.8 Discussion

The scope of this work is to define a comprehensive mathematical framework that allows for the seamless integration of surrogate-based optimization and uncertainty quantification in engineering design problems. The proposed methodology targets the construction of response surfaces of complex stochastic dynamical systems by blending multiple information sources via auto-regressive stochastic modeling. We develop a general setting, under which multi-fidelity in models and multi-fidelity in probability space can coexist in tandem. A computationally efficient framework is developed using multi-level recursive co-kriging and Gaussian Markov random fields, resulting in very robust schemes in two- and three-dimensions.

The capabilities of our implementation were tested for four different cases. In all cases, we were able to successfully exploit cross-correlations within ensembles of surrogate models and efficiently construct an accurate estimate for the response of the given deterministic or stochastic system.

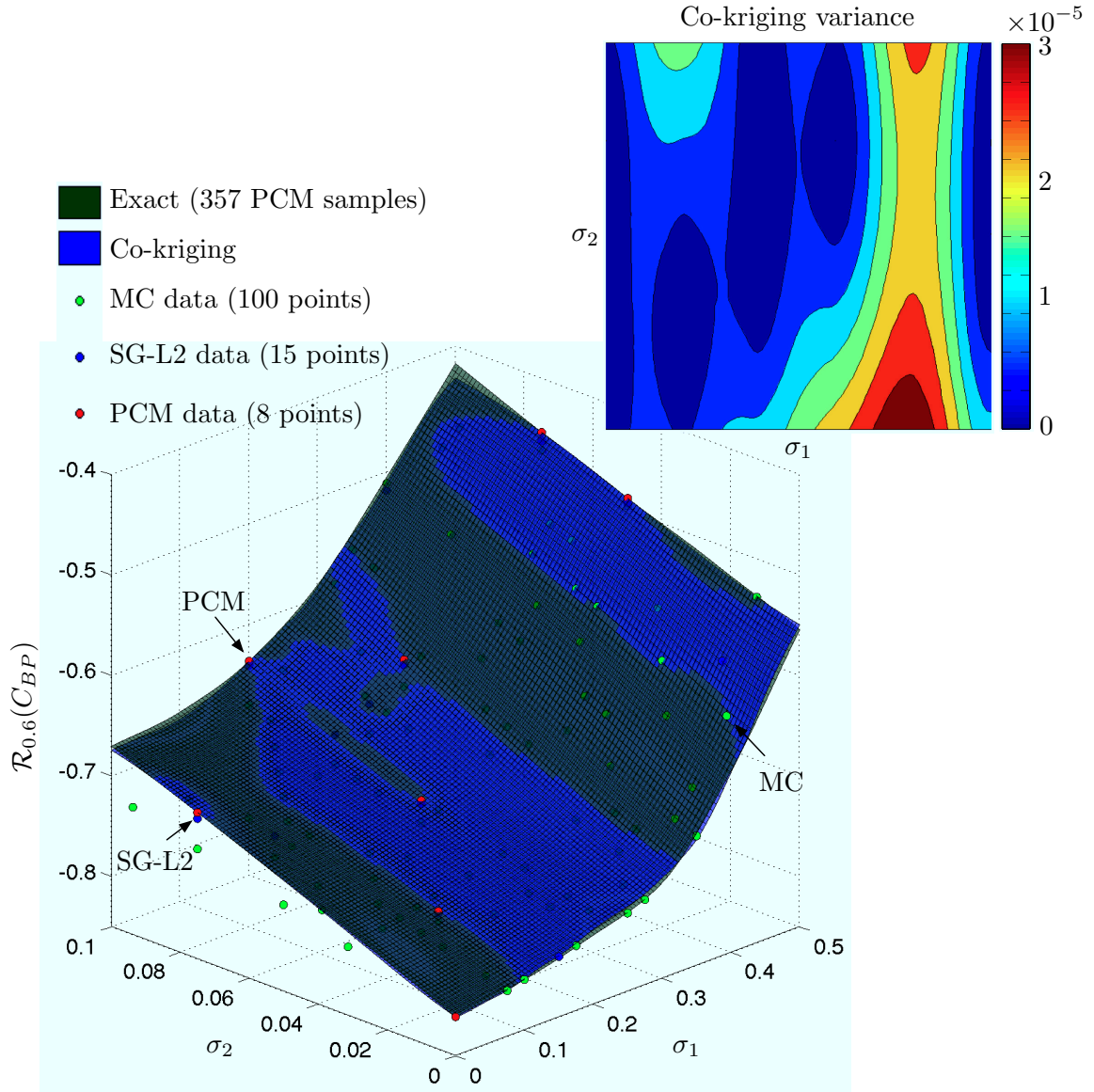


Figure 3.13: *Incompressible stochastic flow past a cylinder*: Response surface of the 0.6-superquantile risk of the base pressure coefficient $\mathcal{R}_{0.6}(C_{BP})$, computed using a 3-level recursive co-kriging scheme, on 100 low-fidelity (MC), 15 intermediate-fidelity (SG-L2), and 5 high-fidelity (PCM) observation points in the (σ_1, σ_2) -plane. The exact surface corresponds to 357 uniformly distributed observations of the highest fidelity model (PCM). The inset plot shows the point-wise variance of the co-kriging predictor.

A limiting factor of our presentation stems from the fact that all demonstrations were confined to two-dimensional problems. Although this significantly contributes towards enhancing the clarity of the presented concepts, we acknowledge that it does not establish a direct link to realistic high-dimensional design problems. At this point we should highlight that one may construct latent variable models through Principal Component Analysis (PCA), non-linear PCA, and projection-pursuit methods (see [Hastie et al. \(2009\)](#)), thus performing a data-driven dimensionality reduction that yields a new set of input coordinates that spans a space of lower dimension.

CHAPTER FOUR

Multi-fidelity Information Fusion Algorithms for High-Dimensional Systems and Massive Data-sets

4.1 Motivation

In pursuit of studying the response of complex engineering and physical systems one has to inevitably face the challenge of interpreting vast amounts of information and quantify the effect of a large number of decision/input variables on the system’s output. Decision making in such data-rich, yet budget-constrained environments, suggests the adoption of an information-driven, machine learning mindset that enables a resourceful way of combining versatile tools, ranging from experiments to stochastic simulations, ultimately shaping new frontiers in data analytics, surrogate-based modeling, design optimization, and beyond.

Ever since the pioneering work of [Sacks et al. \(1989\)](#), the use of surrogate models for the design and analysis of computer experiments has undergone great growth, establishing Gaussian process regression (GPR) as a general and flexible tool for building inexpensive predictive schemes that are capable of emulating the response of complex systems. Furthermore, the use of GPR within auto-regressive stochastic models, such as the widely used scheme put forth by [Kennedy and O’Hagan \(2000\)](#) and the efficient recursive implementation of [Le Gratiet and Garnier \(2014\)](#), allows for exploring spatial cross-correlations between heterogeneous information sources. In Chapter [2](#) we argued that this offers a general platform for developing multi-fidelity information fusion algorithms that simultaneously account for variable fidelity in models (e.g., high-fidelity direct numerical simulations versus low-fidelity empirical formulae) as well as variable fidelity in probability space (e.g., the high-fidelity tensor product multi-element probabilistic collocation of [Foo et al. \(2008\)](#) versus the low-fidelity sparse grid quadratures of [Novak and Ritter \(1996\)](#)). Although this construction is appealing to a wide range of applications, it is mainly limited to low-dimensional input spaces and moderately sized data-sets.

A common strategy for constructing auto-correlation models for GPR in high dimensions is by taking the product of one-dimensional autocorrelation kernels. This typically results to an anisotropic covariance model, which assumes that all dimensions actively interact with each other. However, as the dimensionality is increased, one would hope to find sparsity in the input space, i.e. dimensions with negligible or very weak pairwise interactions. This observation has been widely studied in the literature and has motivated the use of additive models by [Hastie and Tibshirani \(1990\)](#). [Durrande et al. \(2011\)](#) have recently adopted this approach in the context of GPR, advocating versatility in constructing custom auto-correlation kernels that respect the structure in the observed data. This suggests that, having a way to quantify the active interactions in the data, one can tailor an auto-correlation model that closely adheres to those trends. To this end, [Muehlenstaedt et al. \(2012\)](#) have employed functional ANOVA decompositions to compute the degree to which each input dimension, and their pairwise interactions, contribute to the total variability in the observations, and used the corresponding sensitivity indices to construct an undirected graph that provides insight into the structure of possible additive auto-correlation kernels that best suit the available data. Although, this approach is evidently advantageous for scaling GPR to high dimensional problems, it may still suffer from computational tractability issues in presence of big data-sets.

In general, the design of predictive inference schemes in high-dimensions suffers from the well known curse of dimensionality, as the number of points needed to explore the input space in its entirety increases exponentially with the dimension. This implicit need for big data introduces a severe deadlock for scalability in machine learning algorithms as they often involve the repeated inversion of covariance matrices that quantify the spatial cross-correlations in the observations. This defines the so called “big N ” problem - an expression used to characterize the demanding

operational count associated with handling data-sets comprised of N observations ($N > 1000$). The implications of such large data-sets on learning algorithms are well known, leading to a $\mathcal{O}(N^3)$ scaling for implementations based on maximum likelihood estimation. Addressing this challenge has received great attention over the last decades and several methods have been proposed to alleviate the computational cost (see [Dietrich and Newsam \(1997\)](#), [Lindgren et al. \(2011\)](#), [Stein et al. \(2013\)](#)). In Chapter 3 we presented a possible solution path via leveraging the sparsity of GMFR representations. However, in spaces with more than 3 input dimensions this sparsity is lost as the construction of finite element discretizations becomes intractable. Here, we will focus our attention on the frequency-domain learning approach recently put forth by [De Baar et al. \(2013\)](#) that entirely avoids the inversion of covariance matrices at the learning stage and is applicable to a large class of wide-sense stationary auto-correlation models. This essentially enables the development $\mathcal{O}(N)$ algorithms, hence opening the path to predictive inference on massive data-sets.

The aim of this work is to build upon, and further extend, the mathematical framework put forth in Chapter 2, enabling multi-fidelity information fusion for high dimensional systems and in view of massive data-sets. The building block of our construction is GPR and multi-fidelity regression via recursive co-kriging (see [Le Gratiet and Garnier \(2014\)](#)). Motivated by the work of [Muehlenstaedt et al. \(2012\)](#) we pursue scaling to high dimensions by employing hierarchical functional decompositions that reveal structure in the data and inspire a graph-theoretic approach for constructing customized auto-correlation kernels that exploit sparsity in the space of inputs. To this end, we propose a novel, data-driven dimensionality reduction technique based on local projections, and we utilize the Fourier projection-slice theorem (see [Levoy \(1992\)](#)) to decompose the high-dimensional machine learning problem into a series of tractable, low-dimensional problems that can be solved in parallel using $\mathcal{O}(N)$ fast

learning algorithms in the frequency domain.

This chapter employs and further extend the methods presented in Chapter 2, and is structured as follows. In Sec. 4.2.2 we provide an overview of the frequency-domain approach of De Baar et al. (2013) that bypasses the shortcomings of maximum likelihood estimation and enables fast learning from massive data-sets. Subsequently, in Sec. 4.3 we elaborate on kernel design in high dimensions. In particular, we outline the a data-driven hierarchical functional decomposition based on RS-HDMR expansions (see Rabitz et al. (1999)), and describe a graph-theoretic approach inspired by Muehlenstaedt et al. (2012) for tailoring GPR auto-correlation kernels to the data. Moreover, we discuss how to decompose the global high-dimensional learning problem to a series of local solves using a projection-based dimensionality reduction technique in conjunction with the Fourier projection-slice theorem (see Levoy (1992)). In Sec. 4.4 we conclude with a summary of the proposed workflow, underlining key implementation aspects. The capabilities of the proposed methodology are demonstrated through three benchmark problems. First, in Sec. 4.5.1 we employ a multi-fidelity modeling approach for constructing the mean field response of a stochastic flow through a borehole. Second, in Sec. 4.5.2 we present results on the regression of a deterministic function in 100 input dimensions by blending information from three variable fidelity observation models. Lastly, in Sec. 4.5.3 we present an extreme case of performing kriging in 1,000 input dimensions and 10^5 data-points.

4.2 Machine learning in the frequency domain

4.2.1 Bottlenecks of Maximum Likelihood Estimation

Recall the expression for the recursive co-kriging predictor, variance, and likelihood, respectively, presented in Chapter 2 (see Eq. 2.5, 2.6, 2.7)

$$\begin{aligned}\hat{y}_t(\mathbf{x}_t^*) &= \hat{\mu}_t + \hat{\rho}_{t-1}\hat{y}_{t-1}(\mathbf{x}_t^*) + r_t^T(R_t + \hat{\sigma}_{\epsilon_t}^2 I)^{-1}[y_t(\mathbf{x}_t) - \mathbf{1}\hat{\mu}_t - \hat{\rho}_{t-1}\hat{y}_{t-1}(\mathbf{x}_t)], \quad (4.1) \\ s_t^2(\mathbf{x}_t^*) &= \hat{\rho}_{t-1}^2 s_{t-1}^2(\mathbf{x}_t^*) + \hat{\sigma}_t^2 \left[1 - r_t^T(R_t + \hat{\sigma}_{\epsilon_t}^2 I)^{-1}r_t + \frac{[1 - r_t^T(R_t + \hat{\sigma}_{\epsilon_t}^2 I)^{-1}r_t]^2}{\mathbf{1}_t^T(R_t + \hat{\sigma}_{\epsilon_t}^2 I)^{-1}\mathbf{1}_t} \right], \quad (4.2)\end{aligned}$$

$$\begin{aligned}\mathcal{L}(\{\mu_t, \sigma_t^2, \sigma_{\epsilon_t}^2, \rho_{t-1}, \theta_t\}) &= \frac{n}{2} \log(\sigma_t^2) + \frac{1}{2} \log |R_t(\theta_t) + \sigma_{\epsilon_t}^2 I| + \quad (4.3) \\ &+ \frac{1}{2\sigma_t^2} [y_t(\mathbf{x}_t) - \mathbf{1}_t\mu_t - \rho_{t-1}\hat{y}_{t-1}(\mathbf{x}_t)]^T [R_t(\theta_t) + \sigma_{\epsilon_t}^2 I]^{-1} [y_t(\mathbf{x}_t) - \mathbf{1}_t\mu_t - \rho_{t-1}\hat{y}_{t-1}(\mathbf{x}_t)],\end{aligned}$$

The computational cost of calibrating model hyper-parameters through maximum likelihood estimation (MLE) is dominated by the inversion of correlation matrices $(R_t + \sigma_{\epsilon_t}^2 I)^{-1}$ at each iteration needed for the maximization of the likelihood function in Eq. 4.3. The inversion is typically performed using the Cholesky decomposition that scales as $\mathcal{O}(N_t^3)$, leading to a severe bottleneck in the presence of moderately big data-sets. This is typically the case for high-dimensional problems where abundance of data is often required for performing meaningful inference. This pathology is further amplified in cases where the noise variance $\sigma_{\epsilon_t}^2$ is negligible and/or the observed data points are tightly clustered in space. Such cases introduce

ill-conditioning that may well jeopardize the feasibility of the inversion as well as pollute the numerical solution with errors. Moreover, if an anisotropic correlation kernel $\kappa_t(\mathbf{x}_t, \mathbf{x}'_t; \theta_t)$ is assumed, then the vector of correlation lengths θ_t is d -dimensional, leading to an increasingly complex optimization problem as the dimensionality of the input variables $\mathbf{x}_t$ increases. These shortcomings render the learning process intractable for large data-sets, and suggest seeking alternative routes to parameter estimation. Next, we describe a method that bypasses the deficiencies of MLE and enables the development of fast learning algorithms that scale linearly with the data.

4.2.2 Frequency-domain sample variogram fitting

Following the approach of [De Baar et al. \(2013\)](#) we employ the Wiener-Khinchin theorem to fit the auto-correlation function of a wide-sense stationary random field to the power spectrum of the data. The latter contains sufficient information for extracting the second-order statistics that fully describe the Gaussian predictor $Z_t(\mathbf{x})$. Therefore, the model hyper-parameters at each inference level t can be learned in the frequency domain by fitting the Fourier transform of the sample variogram as

$$\min_{\{\sigma_{\epsilon_t}^2, \theta_t\}} \sum_{i=1}^n |\log[\hat{w}_t(\xi_i)]^2 - \log[\hat{a}_t(\xi_i; \sigma_{\epsilon_t}^2, \theta_t)]|^2, \quad (4.4)$$

where ξ is a d -dimensional array of frequencies, $[\hat{w}(\xi)]^2$ is the amplitude of the Fourier coefficients in the modal representation of the data $y_t(\mathbf{x})$, $\hat{a}(\xi; \sigma_{\epsilon_t}^2, \theta_t)$ is the Fourier transform of the auto-correlation function $\{\kappa_t(\mathbf{x}_t, \mathbf{x}'_t; \theta_t) + \sigma_{\epsilon_t}^2 \delta(\|\mathbf{x}_t - \mathbf{x}'_t\|)\}$, with $\delta(\cdot)$ denoting the Dirac delta function, and $\|\cdot\|$ measures distance in an appropriate norm. The Fourier coefficients $\hat{w}(\xi)$ can be efficiently computed with $\mathcal{O}(N_t \log N_t)$ cost using the fast Fourier transform (FFT) for regularly spaced samples, or the

non-uniform fast Fourier transform (NUFFT) (see [Greengard and Lee \(2004\)](#)) for irregularly spaced samples. Moreover, for a wide class of auto-correlation functions, the Fourier transform of $\hat{a}(\xi; \sigma_{\epsilon_t}^2, \theta_t)$ is analytically available, thereby each evaluation of the objective function in the minimization of Eq. 4.4 can be carried out with a linear cost, i.e. $\mathcal{O}(N_t)$. This directly circumvents the limitations of hyper-parameter learning using MLE approaches, namely the cubic scaling associated with inverting dense ill-conditioned correlation matrices, and therefore it enables parameter estimation from massive data-sets.

Although the Wiener-Khinchin theorem relies on the assumption of stationarity, modeling of non-stationary response can be also accommodated by learning a bijective warping of the inputs that removes major non-stationary effects (see [Snoek et al. \(2014\)](#)). This mapping essentially warps the inputs into a jointly stationary space, thus allowing the use of standard wide-sense stationary kernels that enable fast learning in the frequency domain.

A limitation of frequency-domain sample variogram (FSV) fitting is that the summation in Eq. 4.4 is implicitly assumed to take place over all dimensions, i.e. over all N_t^d frequencies in ξ . Although this is tractable for low-dimensional problems, it may easily lead to prohibitive requirements both in terms of memory storage and operation count as the dimensionality increases. In the next section we present a novel methodology for scalable hyper-parameter learning from massive data-sets in high-dimensions.

4.3 Kernel design in high-dimensions

In many physical systems only relatively low-order correlations of the input variables will have an impact on the output (see [Rabitz et al. \(1999\)](#)). In such cases, high-dimensional model representations (HDMR) are proven to dramatically reduce the computational effort in representing the input-output relationships. The general form of the HDMR expansion for a high-dimensional real-valued function $y(\mathbf{x}) = y(x_1, x_2, \dots, x_d)$

$$y(\mathbf{x}) = y_0 + \sum_{1 \leq i \leq d} y_i(x_i) + \sum_{1 \leq i < j \leq d} y_{ij}(x_i, x_j) + \sum_{1 \leq i < j < k \leq d} y_{ijk}(x_i, x_j, x_k) + \dots, \quad (4.5)$$

where y_0 is a constant, $y_i(x_i)$ are functions quantifying the effect of the variable x_i acting independently of all other input variables, $y_{ij}(x_i, x_j)$ represents the cooperative effects of x_i and x_j , while higher order terms reflect the cooperative effects of increasing numbers of variables acting together to impact upon the output of $y(\mathbf{x})$. For a vast range of applications the output is predominantly influenced by two-body interaction terms and rarely terms beyond third-order interactions become significant (see [Rabitz et al. \(1999\)](#)).

Given a set of randomly sampled scattered observations, one can employ the RS-HDMR construction (see [Li et al. \(2002\)](#)) to represent the data according to the hierarchical functional decomposition of Eq. 4.5. To this end, we first rescale the input variables $\mathbf{x}$ such that $0 \leq x_i \leq 1$, $i = 1, \dots, d$. Therefore, the support of the output $y(\mathbf{x})$ is the unit hypercube $\mathcal{K}^d$, and the RS-HDMR component functions take the form

$$y_0 = \int_{\mathcal{K}^d} y(\mathbf{x}) d\mathbf{x} \quad (4.6)$$

$$y_i(x_i) = \int_{\mathcal{K}^{d-1}} y(\mathbf{x}) d\mathbf{x}^{(i)} - y_0 \quad (4.7)$$

$$y_{ij}(x_i, x_j) = \int_{\mathcal{K}^{d-2}} y(\mathbf{x}) d\mathbf{x}^{(ij)} - y_i(x_i) - y_j(x_j) - y_0 \quad (4.8)$$

$$\vdots$$

where $d\mathbf{x}^{(i)}$ and $d\mathbf{x}^{(ij)}$ correspond to the product $dx_1 dx_2 \dots dx_d$, excluding dx_i and $dx_i dx_j$, respectively. This structure suggests that a Monte Carlo, or quasi-Monte Carlo, or multi-level Monte Carlo, or multi-element probabilistic collocation integration strategies can be employed to directly determine the component functions from data (see [Li et al. \(2002\)](#), [Foo and Karniadakis \(2010\)](#)). However, the required number of random samples for achieving reasonable accuracy is known to increase exponentially with the order of the RS-HDMR component functions (see [Li et al. \(2002\)](#)). To this end, one may reduce the sampling effort significantly by approximating the component functions using tensor product expansions of orthonormal polynomials (see [Li et al. \(2002\)](#))

$$y_i(x_i) \approx \sum_{r=1}^k \alpha_r^i \phi_r(x_i) \quad (4.9)$$

$$y_{ij}(x_i, x_j) \approx \sum_{p=1}^l \sum_{q=1}^{l'} \beta_{pq}^{ij} \phi_p(x_i) \phi_q(x_j) \quad (4.10)$$

$$\vdots$$

where α_r^i and β_{pq}^{ij} are unknown expansion coefficients determined from data, and $\phi_r(x)$ are basis functions of order r . For a set of orthonormal basis functions, the

unknown coefficients can be directly determined from the N_t data-points at each inference level via Monte Carlo integration

$$\alpha_r^i = \int_{\mathcal{K}^d} y(\mathbf{x}) \phi_r(x_i) d\mathbf{x} \approx \frac{1}{N_t} \sum_{s=1}^{N_t} y(\mathbf{x}^{(s)}) \phi_r(x_i^{(s)}) \quad (4.11)$$

$$\beta_{pq}^{ij} = \int_{\mathcal{K}^d} y(\mathbf{x}) \phi_p(x_i) \phi_q(x_j) d\mathbf{x} \approx \frac{1}{N_t} \sum_{s=1}^{N_t} y(\mathbf{x}^{(s)}) \phi_p(x_i^{(s)}) \phi_q(x_j^{(s)}) \quad (4.12)$$

$\vdots$

Moreover, the RS-HDMR construction allows for a straightforward quantification of the sensitivity D_i of the output $y(\mathbf{x})$ on each input variable x_i , their pairwise interactions D_{ij} , as well as higher-order interactions:

$$D_i = \int_0^1 y_i^2(x_i) dx_i \approx \int_0^1 \left[\sum_{r=1}^{k_i} \alpha_r^i \phi_r(x_i) \right]^2 dx_i = \sum_{r=1}^{k_i} (\alpha_r^i)^2 \quad (4.13)$$

$$D_{ij} = \int_0^1 \int_0^1 y_{ij}^2(x_i, x_j) dx_i dx_j \approx \int_0^1 \int_0^1 \left[\sum_{p=1}^{l_i} \sum_{q=1}^{l'_j} \beta_{pq}^{ij} \phi_p(x_i) \phi_q(x_j) \right]^2 dx_i dx_j = \sum_{p=1}^{l_i} \sum_{q=1}^{l'_j} (\beta_{pq}^{ij})^2 \quad (4.14)$$

$\vdots$

The sensitivity indices identify active interactions in high-dimensional data-sets. This valuable information can guide the design of correlation kernels that are tailored to the given data-set, respecting all significant input-output interactions. To this end, we employ a graph-theoretic approach in which custom correlation kernels can

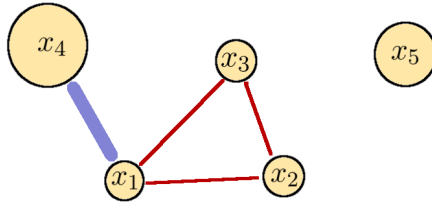


Figure 4.1: *Kernel design in high-dimensions*: Sketch of the undirected graph defined by the Sobol sensitivity indices of 5-dimensional function $y(x_1, x_2, x_3, x_4, x_5)$. Here we have one third-order interaction, four second-order interactions and five first-order terms. The size of the lines and the circles indicate the magnitude of the sensitivities. The corresponding maximal cliques are $\mathcal{C}_1 = \{x_1, x_2, x_3\}$, $\mathcal{C}_2 = \{x_1, x_4\}$, $\mathcal{C}_3 = \{x_5\}$.

be constructed as an *additive* composition of kernels that describe cross-correlations within each one of the maximal cliques of the undirected graph defined by the HDMR sensitivity indices. The first step towards this construction involves assembling the undirected graph $G = (V, E)$ of the computed sensitivity indices, where first-order sensitivities correspond to vertices V , while sensitivity indices of higher-order interactions define edges E . Once the undirected graph is available, a clique, $\mathcal{C}$, can be identified as a subset of the vertices, $\mathcal{C} \subseteq V$, such that every two distinct vertices are adjacent. This is equivalent to the condition that the subgraph of G induced by $\mathcal{C}$ is complete. A maximal clique is a clique that cannot be extended by including one more adjacent vertex, that is, a clique which does not exist exclusively within the vertex set of a larger clique (see Fig. 4.1). Maximal cliques can be efficiently identified from the graph of sensitivity indices using the Bron-Kerbosch algorithm with both pivoting and degeneracy re-ordering (see [Bron and Kerbosch \(1973\)](#)).

This procedure reveals the extent to which the observed data encodes an additive structure. The key idea here is to exploit this structure in order to effectively decompose the high-dimensional learning problem in a sequence of lower dimensional tasks, where estimation of model hyper-parameters can take place independently within the support of each one of the maximal cliques. To this end, recall that fitting the sample

variogram in the frequency domain becomes intractable in high dimensions as the objective functional in Eq. 4.4 involves storing and computing N^d entries. However, we can utilize the RS-HDMR representation in order to exploit the structure encoded in the maximal cliques, and efficiently perform FSV fitting locally for each maximal clique. This can be done by constructing an additive auto-correlation kernel that fully respects the active interactions within each maximal clique

$$\kappa(\mathbf{x}, \mathbf{x}'; \theta) = \sum_{q=1}^{N_C} \kappa_q(\mathbf{x}_q, \mathbf{x}'_q; \theta_q), \quad (4.15)$$

where N_C is the total number of maximal cliques at each recursive co-kriging level. Our goal now is to estimate the hyper-parameters θ_q by fitting the Fourier transform of each auto-correlation kernel $\kappa_q(\mathbf{x}_q, \mathbf{x}'_q; \theta_q)$ to the power spectrum of the data. In order to do so, we first have to identify the contribution of each maximal clique to the power spectrum of the d -dimensional data-set. To this end, the RS-HDMR expansion terms can be utilized to project data onto the sub-manifold defined by each maximal clique as

$$\mathcal{P}_q y(x) = y_0 + \sum_{i \in \mathcal{C}_q} y_i(x_i) + \sum_{i,j \in \mathcal{C}_q} y_{ij}(x_i, x_j) + \sum_{i,j,k \in \mathcal{C}_q} y_{ijk}(x_i, x_j, x_k) + \dots, \quad 1 \leq q \leq N_C, \quad (4.16)$$

where $\mathcal{C}_q$ is an index set listing all active dimensions contained in the q^{th} maximal clique, and the projection operator $\mathcal{P}_q$ is a multivariate integral with respect to all input variables x_q that do not appear in $\mathcal{C}_q$. Then, by assuming a wide-sense stationary covariance kernel $\kappa_q(\mathbf{x}_q, \mathbf{x}'_q; \theta_q)$ in each maximal clique, we can employ the FSV learning algorithm to estimate θ_q by fitting the power spectrum of the clique-projected data, which typically live in a sub-manifold of dimension much lower than d . This approach is motivated by the Fourier projection-slice theorem (see [Levoy \(1992\)](#)), which formalizes the equivalence between taking the Fourier transform of a

projection versus taking a slice of the full high-dimensional spectrum. In operator terms, this translates to

$$\mathcal{F}_m \mathcal{P}_m = \mathcal{S}_m \mathcal{F}_d, \quad (4.17)$$

where $\mathcal{F}_m$ is an m -dimensional Fourier transform, $\mathcal{P}_m$ is an operator projecting d -dimensional data onto an m -dimensional sub-manifold, and $\mathcal{S}_m$ is an m -dimensional slice operator that returns an m -dimensional linear sub-manifold through the origin in the Fourier space, which is parallel to the projection sub-manifold (see [Levoy \(1992\)](#)). In our case, the q^{th} sub-manifold lives in the support of the active dimensions contained in the maximal clique $\mathcal{C}_q$, and the projection $\mathcal{P}_q$ can be directly computed by evaluating Eq. 4.16 on a regular grid that discretizes the hypercube defined by $\mathcal{H}_q = \{x_i, i \in \mathcal{C}_q | 0 \leq x_i \leq 1\}$. The main advantage here is that for high-dimensional cases that admit an additive RS-HDMR representation, the dimension of the q^{th} sub-manifold is $m = \text{card}\{\mathcal{C}_q\} \ll d$. Hence, the optimal hyper-parameters $\hat{\theta}_i$ defining the auto-correlation kernel in each clique can be estimated very efficiently using the FSV fitting algorithm. Finally, summing up all clique contributions in Eq. 4.15 we can obtain a global correlation kernel that according to the Fourier projection-slice theorem captures the power spectrum of the original high-dimensional observations. This allows us to fit Gaussian random field models to big data in high-dimensions by using $\mathcal{O}(N)$ learning algorithms!

4.4 Implementation aspects

Here we provide an overview of the workflow and discuss some key implementation aspects.

Step 1: Starting from a set of available scattered observations $y_t(\mathbf{x})$ at the inference level $1 \leq t \leq s$, our first task is to compute the RS-HDMR representation. To this end, we adopt the approach of [Ziehn and Tomlin \(2009\)](#) that employs an orthonormal basis of shifted Legendre polynomials (up to order 15), using adaptive criteria for the optimal selection of the polynomial order that approximates each component function, and variance reduction techniques that enhance the accuracy of the RS-HDMR representation when only a limited number of samples is available. For all cases considered, an RS-HDMR expansion with up to second-order interaction terms was sufficient to capture more than 95% of the variance in the observations.

Step 2: Once the RS-HDMR representation is computed, we invoke the Bron-Kerbosch algorithm (see [Bron and Kerbosch \(1973\)](#)) to identify all maximal cliques in the undirected graph of sensitivity indices. This guides the construction of an additive auto-correlation kernel (see Eq. 4.15) that is comprised of all clique-wise contributions. Throughout all recursive co-kriging inference levels we have assumed an anisotropic product Gaussian auto-correlation function for all corresponding maximal cliques, $1 \leq q \leq N_C$.

$$\kappa_q(\mathbf{x}_q, \mathbf{x}'_q; \theta_q) = \prod_{i=1}^m e^{-\frac{|\mathbf{x}_i - \mathbf{x}'_i|^2}{2\theta_i}} \quad (4.18)$$

where $m = \mathbf{card}\{\mathcal{C}_q\}$ is the number of dimensions contained in the q^{th} clique, and θ_i is a correlation length hyper-parameter along the i^{th} dimension. In this case, the Fourier transform of the auto-correlation function in Eq. 4.4 is available analytically

$$\hat{a}(\xi_q; \sigma_{\epsilon_q}^2, \theta_q) = \sigma_{\epsilon_q}^2 + (2\pi)^{\frac{m}{2}} \theta_q e^{-2 \sum_{i=1}^m (\pi \theta_i \xi_i)^2}. \quad (4.19)$$

Step 3: The next step involves learning the hyper-parameters $\{\sigma_{\epsilon_q}^2, \theta_q\}$ for each maximal clique at inference level t by fitting the power spectrum of the clique-projected data (see Eqs. 4.4, 4.16). To this end, the data is projected on a regular grid with 128 points along each clique dimension. This corresponds to using 128 Fourier modes for resolving the variability of the shifted Legendre basis functions in the frequency domain. Note that the learning task is directly amenable to parallelization as it can be performed independently for each maximal clique. Once the optimal values of $\{\hat{\sigma}_{\epsilon_q}^2, \hat{\theta}_q\}$ are known, we can construct the global correlation matrix (see Eq. 4.15) and factorize it using the Cholesky decomposition. Then, the optimal values of $\{\hat{\mu}_t, \hat{\rho}_{t-1}, \hat{\sigma}_t^2\}$ can be obtained from Eqs. 2.8, 2.9, where $(R_t + \sigma_{\epsilon_t}^2 I)^{-1}$ is computed via back-substitution of the Cholesky factors.

Step 4: Finally, given a set of prediction points $\mathbf{x}^*$ we can employ Eqs. 4.1, 4.2 to evaluate the predictor $\hat{y}_t(\mathbf{x}_t^*)$ and variance $s_t^2(\mathbf{x}_t^*)$. This task is also trivially parallelizable as predictions at different points in $\mathbf{x}_t^*$ can be performed independently of each other.

4.5 Results

4.5.1 Borehole function

The first benchmark case involves a set of multi-fidelity functions that simulate water flow through a borehole. The functions correspond to two fidelity levels in physical space and depend on 8 input parameters and 4 random variables. The borehole function models water flow through a borehole, returning the water flow rate in $m^3/year$ as a function of 8 input parameters (see [Surjanovic and Bingham](#)).

Here, we assume that high fidelity observations can be generated using the following expression

$$f_h(\mathbf{x}) = \frac{2\pi T_u(H_u - H_l)}{\log(r/r_w) \left(1 + \frac{2LT_u}{\log(r/r_w)r_w^2 K_w} + \frac{T_u}{T_l}\right)}, \quad (4.20)$$

where $\mathbf{x} = [r_w, r, T_u, H_u, T_l, H_l, L, K_w]$, $r_w \in [0.05, 0.15]$ is the radius of the borehole (in m), $r \in [100, 50000]$ is the radius of influence (in m), $T_u \in [63070, 115600]$ is the transitivity of the upper aquifer (in m^2/year), $H_u \in [990, 1110]$ is the potentiometric head of the upper aquifer (in m), $T_l \in [63.1, 116]$ is the transmissivity of the lower aquifer (in m^2/year), $H_l \in [700, 820]$ is the potentiometric head of the lower aquifer (in m), $L \in [1120, 1680]$ is the length of the borehole (in m), and $K_w \in [9855, 12045]$ is the hydraulic conductivity of the borehole (in m/year).

We also assume that observations generated by Eq. 4.20 are perturbed by a noise term $\eta(\mathbf{z})$ expressed as a function of 4 standard normal random variables, $z_i \sim \mathcal{N}(0, 1)$, $i = 1, \dots, 4$

$$\eta(z_1, z_2, z_3, z_4) = z_1 \sin^2[(2z_2 + z_3)\pi] - \cos^2(z_4\pi), \quad (4.21)$$

thus generating stochastic high-fidelity data of the form

$$y_h(\mathbf{x}; \mathbf{z}) = f_h(\mathbf{x})[1 + 0.2\eta(\mathbf{z})], \quad (4.22)$$

where we have assumed a noise amplitude of 20%. Similarly, we can obtain stochastic low-fidelity data $y_l(\mathbf{x}; \mathbf{z})$ by substituting f_h in Eq. 4.22 with a lower fidelity model given by

$$f_l(\mathbf{x}) = \frac{5T_u(H_u - H_l)}{\log(r/r_w) \left(1.5 + \frac{2LT_u}{\log(r/r_w)r_w^2 K_w} + \frac{T_u}{T_l} \right)}. \quad (4.23)$$

Our goal here is to employ the proposed multi-fidelity information fusion framework to construct the response surface of the 8-dimensional mean field $S(\mathbf{x}) = \mathbb{E}[Y(\mathbf{x}; \mathbf{z})]$, using the observation model of Eq. 3.51 given observations $\{y_h(\mathbf{x}; \mathbf{z}), y_l(\mathbf{x}; \mathbf{z})\}$. In order to approximate the expectation operator we employ two methods of different fidelity in probability space. To this end, we choose the high-fidelity probabilistic method to be a Smolyak sparse grid level-5 quadrature rule (SG-L5) (see [Novak and Ritter \(1996\)](#)) that discretizes the 4-dimensional parameter space using 4,994 quadrature points. Similarly, the low-fidelity probabilistic quadrature is a coarser sparse grid level-2 (SG-L2) rule with just 57 quadrature points.

Consequently, the available multi-fidelity information sources consist of two models in physical space $(y_h(\mathbf{x}), y_l(\mathbf{x}))$, and two models in probability space (SG-L5, SG-L2). This results to a family of response surfaces S_{ij} that can be organized as

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} \mathbb{E}_{SG-L2}[y_l(\mathbf{x}; \mathbf{z})] & \mathbb{E}_{SG-L5}[y_l(\mathbf{x}; \mathbf{z})] \\ \mathbb{E}_{SG-L2}[y_h(\mathbf{x}; \mathbf{z})] & \mathbb{E}_{SG-L5}[y_h(\mathbf{x}; \mathbf{z})] \end{pmatrix} \quad (4.24)$$

Here, we employ a 4-level recursive co-kriging scheme (see Sec. 2.3.2), traversing the available models and data in the order $S_{11} \rightarrow S_{12} \rightarrow S_{21} \rightarrow S_{22}$. Following the steps outlined in Sec. 4.4, first we compute the RS-HDMR representation of the data considering a set of randomly sampled training points $[1024, 512, 128, 32]$ that corresponds to each one of the four observation models with increasing fidelity. Figure 4.2 shows the resulting undirected graph of sensitivity indices that characterizes the active interactions in the data. Note that for all cases, an HDMR expansion

with up to second-order interaction terms was sufficient to capture more than 98% of the variance in the observations of each model. Using the undirected graph we can identify the maximal cliques, leading to the following additive decomposition of the Gaussian auto-corellation kernel

$$\kappa(\mathbf{x}, \mathbf{x}'; \theta) = \prod_{i \in \{1,4\}} e^{-\frac{|\mathbf{x}_i - \mathbf{x}'_i|^2}{2\theta_i}} + \prod_{i \in \{1,6\}} e^{-\frac{|\mathbf{x}_i - \mathbf{x}'_i|^2}{2\theta_i}} + \quad (4.25)$$

$$+ \prod_{i \in \{1,7\}} e^{-\frac{|\mathbf{x}_i - \mathbf{x}'_i|^2}{2\theta_i}} + e^{-\frac{|\mathbf{x}_2 - \mathbf{x}'_2|^2}{2\theta_2}} + e^{-\frac{|\mathbf{x}_3 - \mathbf{x}'_3|^2}{2\theta_3}} + \quad (4.26)$$

$$+ e^{-\frac{|\mathbf{x}_5 - \mathbf{x}'_5|^2}{2\theta_5}} + e^{-\frac{|\mathbf{x}_8 - \mathbf{x}'_8|^2}{2\theta_8}}, \quad (4.27)$$

corresponding to 7 maximal cliques $\mathcal{C}_1 = \{1, 4\}$, $\mathcal{C}_2 = \{1, 6\}$, $\mathcal{C}_3 = \{1, 7\}$, $\mathcal{C}_4 = \{2\}$, $\mathcal{C}_5 = \{3\}$, $\mathcal{C}_6 = \{5\}$, and $\mathcal{C}_7 = \{8\}$. Next, we utilize the available observations to calibrate the auto-correlation hyper-parameters θ by solving $N_{\mathcal{C}} = 7$ FSV learning problems, involving fitting the power spectrum of the available data projected onto each one of the maximal cliques. Finally, we sum up all the clique-wise contributions to obtain the global auto-correlation kernel $\kappa(\mathbf{x}, \mathbf{x}'; \theta)$ which is used to construct the correlation matrix R_t at each co-kriging level, $t = 1, \dots, 4$. Finally, R_t is factorized once using the Cholesky decomposition leading to an optimal set of $\{\hat{\mu}_t, \hat{\rho}_{t-1}, \hat{\sigma}_t^2\}$ (see Eqs. 2.8, 2.9), and thus enabling the computation of the predictions and variance according to Eqs. 4.1, 4.2.

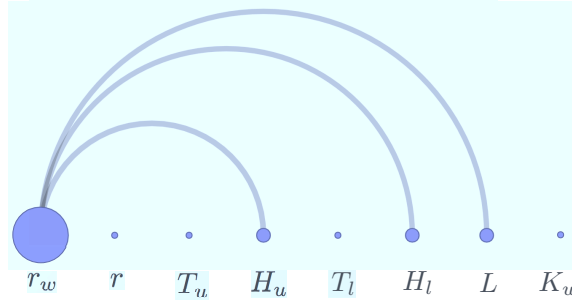


Figure 4.2: *Borehole function*: Sketch of the undirected graph of the RS-HDMR sensitivity indices generated using 1,024 observations of $S_{11}(\mathbf{x})$. The first-order sensitivity indices are $S_i = [0.837, 0.0, 0.0, 0.042, 0.0, 0.041, 0.041, 0.0]$, while the non-zero pair-wise sensitivities are $S_{14} = 0.011$, $S_{16} = 0.011$, and $S_{14} = 0.010$, capturing 99.7% of the total variance in the observations.

Figure 4.3 presents a qualitative assessment of the accuracy of the co-kriging predictor of the quantity of interest $\mathbb{E}[y]$. In particular, we have considered a set of 2,000 uniformly distributed points $\mathbf{x}^*$ in the 8-dimensional input space, where we test the accuracy of the co-kriging predictor $\hat{y}_t(\mathbf{x}^*)$ at each level versus an “exact” solution constructed computing $\mathbb{E}[y]$ using 10^6 Monte Carlo samples of the highest fidelity observation model $f_h(\mathbf{x}^*)$. Figure 4.3(a) shows a density plot of the frequency distribution of the exact solution $\mathbb{E}[y]$, versus the estimated $\mathbb{E}[\hat{y}]$ resulting from the co-kriging predictors at each level. Moreover, we have also included the non-probabilistic prediction of the RS-HDMR representation (Eq. 4.5). The output of the co-kriging predictors has been also plotted in Figure 4.3(b) in a scatter plot format. Evidently, the response surface of the mean field is captured remarkably well by just using 32 observations of the highest fidelity model S_{22} , supplemented by a number of inaccurate but mutually correlated low-fidelity observations from (S_{11}, S_{12}, S_{21}) . Lastly, note that the total elapsed CPU time for performing both the learning and the prediction steps was about 5 minutes on a single core of Intel Xeon E5607@2.27GHz, requiring about 3 megabytes of memory.

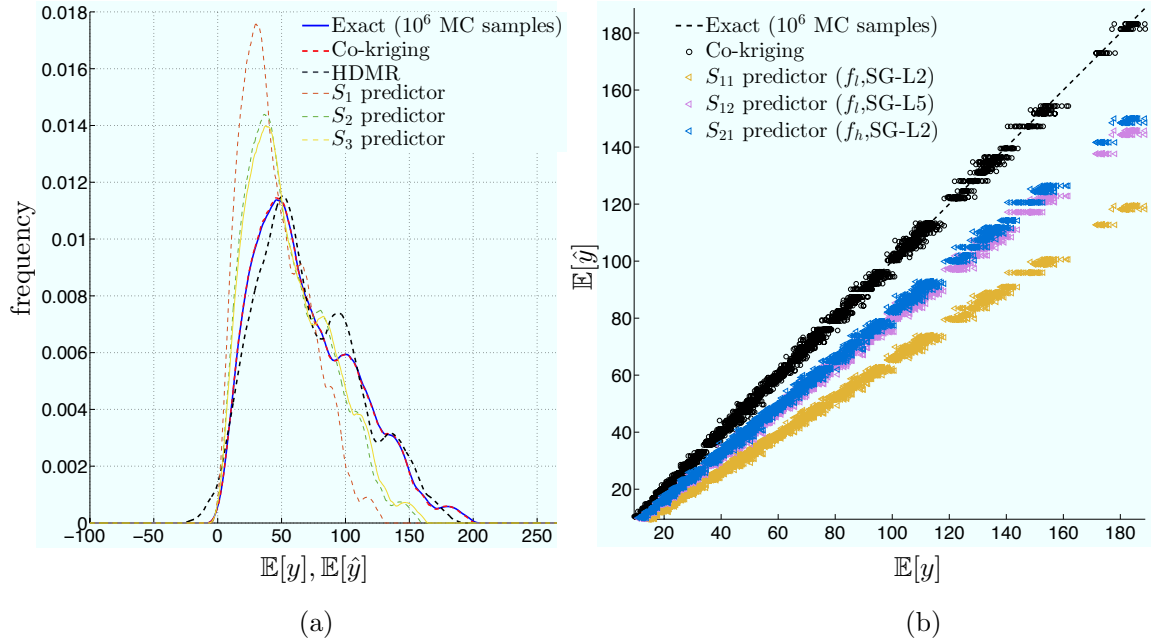


Figure 4.3: *Borehole function*: *Left*: Density plot of the frequency distribution of the exact solution $\mathbb{E}[y]$ (blue solid line), versus the estimated $\mathbb{E}[\hat{y}]$ (dashed lines) resulting from co-kriging and the HDMR representation. The red dashed line corresponds to the final co-kriging predictor accounting for information fusion along the path $S_{11} \rightarrow S_{12} \rightarrow S_{21} \rightarrow S_{22}$. *Right*: Scatter plot of the exact solution $\mathbb{E}[y]$ (black dashed line), versus the co-kriging predictor $\mathbb{E}[\hat{y}]$ at each level (circles) at 2,000 randomly sampled test locations. The black circles correspond to the final co-kriging predictor accounting for information fusion along the path $S_{11} \rightarrow S_{12} \rightarrow S_{21} \rightarrow S_{22}$.

4.5.2 Trid function in 100 dimensions

This example involves the multi-fidelity regression of the deterministic Trid function in 100 input dimensions. Here we employ three models of variable fidelity in physical space, all defined in the hypercube $[-100^2, 100^2]^{100}$ and taking the form

$$y_3(\mathbf{x}) = \sum_{i=1}^{100} (x_i - 1)^2 - \sum_{i=2}^{100} x_i x_{i-1} \quad (4.28)$$

$$y_2(\mathbf{x}) = 0.6y_3(\mathbf{x}) + 2\|\cos(\mathbf{x}\pi)\|_2 - 4 \quad (4.29)$$

$$y_1(\mathbf{x}) = 0.3y_3(\mathbf{x}) - 3\|\sin(\mathbf{x}\pi)\|_2 + 6, \quad (4.30)$$

where lower fidelity models are obtained by perturbing the exact representation encoded by $y_3(\mathbf{x})$. In particular, we consider $y_1(\mathbf{x})$ to be the lowest fidelity observation model, while $y_2(\mathbf{x})$ returns observations of intermediate fidelity.

Starting from a training set containing [8196, 2048, 256] randomly sampled observations (lowest to highest fidelity) the RS-HDMR representation is constructed and the corresponding undirected graph of sensitivity indices is presented in Fig. 4.4. For all fidelity levels, an RS-HDMR expansion containing up to second order terms can capture more than 98% of the variance in the observations. The sensitivity graph is in agreement with Eq. 4.28, which clearly states that observations should contain only pairwise interactions, thus yielding a set of 99 two-dimensional maximal cliques, and suggesting an auto-correlation kernel of the form

$$\kappa(\mathbf{x}, \mathbf{x}'; \theta) = \sum_{i=1}^{99} \prod_{j=i}^{i+1} e^{-\frac{|\mathbf{x}_j - \mathbf{x}'_j|^2}{2\theta_j}}. \quad (4.31)$$

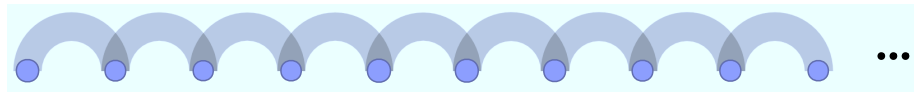


Figure 4.4: *Trid function*: Sketch of the undirected graph of the RS-HDMR sensitivity indices generated using 8,196 observations of $y_1(\mathbf{x})$. The structure is in agreement with the definition of the Trid function in Eq. 4.28, suggesting the presence of only pairwise interactions that lead to 99 maximal cliques, each containing two active dimensions.

Following the workflow presented in Sec. 4.4 we can solve the clique-wise learning problems to calibrate the correlation length hyper-parameters to the power spectrum of the clique-projected data, and then utilize the co-kriging predictive distribution (Eqs. 4.1, 4.2) to estimate the response generated by fusing information along the path of $y_1 \rightarrow y_2 \rightarrow y_3$. Figure 4.5 summarizes the result of the proposed methodology by comparing the co-kriging and HDMR predictors to the exact function values obtained at a test set of 2,000 test points. Note that here, due to the high-dimensional setting and the limited number of samples, it is not meaningful to randomly probe the solution the whole input space, as the predictor cannot be expected to be accurate at locations far away from the training points. To this end, in order to provide a meaningful assessment of accuracy, we have chosen the test points by uniformly sampling within a unit hyper-sphere centered at some of the observation points. Of course, as the radius of the hyper-sphere is increased we expect a gradual deterioration of the accuracy in the predictions, although our tests indicate that accuracy is preserved for a radius equal to the maximal correlation length in the data. Both the density (Fig. 4.5(a)) and scatter plot (Fig. 4.5(b)) indicate that the proposed recursive co-kriging approach can accurately represent the high-dimensional response surface by exploring spatial cross-correlations between few high-fidelity observations and a set of inaccurate, lower fidelity samples. First, in Figs. 4.5(a) we observe that the RS-HDMR predictor returns erroneous results; a sign that more samples or more aggressive variance reduction techniques are needed in the computation of the RS-HDMR components. On the other hand, even when the RS-HDMR representation returns non-credible predictions, the resulting co-kriging scheme with the additive clique-based auto-correlation kernel retains high accuracy at the neighborhood of observations. This observation reveals a degree of robustness of the proposed kriging/co-kriging schemes with respect to misspecification of the RS-HDMR expansion.

The total elapsed CPU time for this benchmark problem was about 30 minutes on a single core of Intel Xeon E5607@2.27GHz, requiring about 20 megabytes of memory.

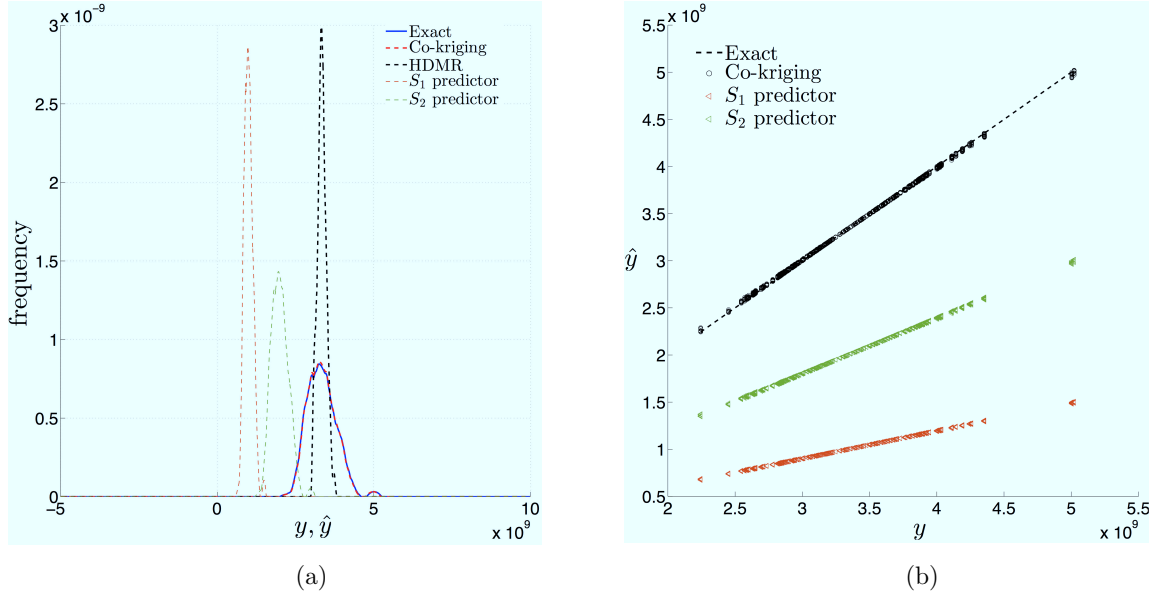


Figure 4.5: *Trid* function: *Left*: Density plot of the of the frequency distribution of the exact solution $y = y_3(\mathbf{x}^*)$ (blue solid line), versus the estimated $\hat{y}(\mathbf{x}^*)$ (dashed lines) resulting from co-kriging and the HDMR representation. The red dashed line corresponds to the final co-kriging predictor accounting for information fusion along the path $y_1 \rightarrow y_2 \rightarrow y_3$. *Right*: Scatter plot of the of the exact solution $y = y_3(\mathbf{x}^*)$ (black dashed line), versus the co-kriging predictor $\hat{y}(\mathbf{x}^*)$ at each level (circles) at 2,000 test locations, randomly sampled within unit hyper-spheres centered at observations. The black circles correspond to the final co-kriging predictor accounting for information fusion along the path $y_1 \rightarrow y_2 \rightarrow y_3$.

4.5.3 Sobol function in 1,000 dimensions

In this last example we have considered an extreme case involving regression of the Sobol function in 1,000 dimensions. The Sobol function is a tensor product function that is routinely used as a benchmark problem in sensitivity analysis. Here, we consider the input space defined by the unit hypercube $[0, 1]^{1000}$, and assume just a single level of fidelity represented by

$$y(\mathbf{x}) = \prod_{i=1}^{1000} \frac{|4x_i - 2| + a_i}{1 + a_i}, \quad (4.32)$$

where $a_i = i^2$, and for each index i , a lower value of a_i indicates a higher importance of the input variable x_i . Although this tensor product form assumes that all dimensions are actively interacting with each other, [Zhang et al. \(2012\)](#) have demonstrated that for this particular choice of a_i , the effective dimensionality of the Sobol function is much lower than 1,000, and an additive representation with up to second order interaction terms is capable of capturing more than 97% of the variance.

To demonstrate the capability of the proposed framework to simultaneously handle high dimensions and massive data-sets, we have chosen here a training set containing 10^5 randomly sampled observations. Moreover, we have kept only first-order terms in the RS-HDMR representation to highlight the extent to which a purely additive decomposition can capture the tensor product response. As a consequence of this assumption, the undirected graph of sensitivity indices reveals a structure of 1,000 maximal cliques, each one containing a single active dimension, thus suggesting an additive auto-correlation kernel of the form

$$\kappa(\mathbf{x}, \mathbf{x}'; \theta) = \sum_{i=1}^{1000} e^{-\frac{|\mathbf{x}_i - \mathbf{x}'_i|^2}{2\theta_i}}. \quad (4.33)$$

Once more, we compute the distribution of the kriging predictor following the steps outlined in Sec. 4.4. The result is summarized in Figs. 4.6(a),(b), in form of a density and a scatter plot, respectively, that show the accuracy of the kriging and RS-HDMR predictors versus the exact solution given by Eq. 4.32. Again, we have considered 2,000 test points that are uniformly sampled within the unit hypercube. First, in Fig. 4.6(a) we observe that the RS-HDMR predictor returns a fairly

accurate distribution, confirming the fact that the response admits a purely additive representation. The quality of the RS-HDMR predictor in this case is also enhanced by the large number of samples that guided an accurate computation of the first-order component functions in Eq. 4.5. Figures 4.6(a),(b), also characterize the behavior of a kriging scheme with the additive Gaussian auto-correlation model of Eq. 4.33. There, we observe that the kriging predictive distribution provides an excellent fit to the exact solution, accurately reproducing the distribution density and returning an mean square error of 0.097 (see Figs. 4.6(b)).

The computational imprint of this analysis was 70 minutes of CPU time, requiring about 800 megabytes of memory. Table 4.1 presents a detailed breakdown of the computational cost with respect to the steps outlined in Sec. 4.4, on a single core of Intel Xeon E5607@2.27GHz. Note, that the reported values correspond to a serial implementation, although both Steps 1, 4 are directly amenable to parallelization, potentially leading to a great reduction in the CPU time. Due to the high dimensionality and the large number of observations, the cost is dominated by the computation of the RS-HDMR component functions (69%), followed by the prediction step (25%). In contrast, for problems of lower dimensionality and moderately sized data-sets (e.g., the Borehole function case of Sec. 4.5.1) the cost is typically attributed to learning the hyper-parameters within each maximal clique (see Step 3 in Sec. 4.4).

Step	Time elapsed (sec)	% of total time
1	2,948	69%
2	0.097	$\ll 1\%$
3	237	6%
4	1,073	25%

Table 4.1: *Sobol function*: Breakdown of the total computational cost with respect to the implementation steps outlined in Sec. 4.4.

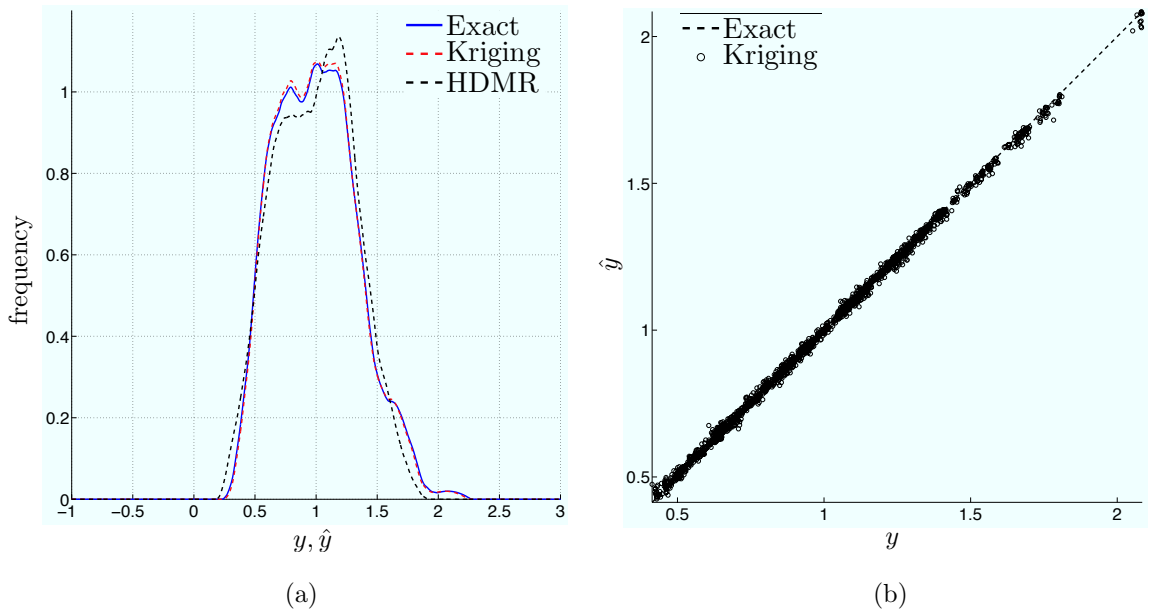


Figure 4.6: *Sobol function*: *Left*: Density plot of the frequency distribution of the exact solution $y(\mathbf{x}^*)$ (*blue* solid line), versus the estimated $\hat{y}(\mathbf{x}^*)$ (dashed lines) resulting from kriging and the HDMR representation. *Right*: Scatter plot of the exact solution $y(\mathbf{x}^*)$ (*black* dashed line), versus the kriging predictor $\hat{y}(\mathbf{x}^*)$ (*black* circles) at 2,000 test locations, randomly sampled within unit hyper-spheres centered at observations.

4.6 Discussion

The scope of this chapter is to propose a tractable framework for multi-fidelity information fusion in high dimensions and in presence of massive data-sets. This framework is based on Gaussian process regression and auto-regressive stochastic schemes, further extending the methodology presented in Chapter 2 that targets the seamless integration of surrogate-based optimization and uncertainty quantification for high-dimensional engineering design problems, and beyond. Specifically, the proposed algorithms allow for the efficient construction of response surfaces of complex stochastic dynamical systems by blending variable information sources through Gaussian process regression and auto-regressive stochastic schemes. Scaling to high dimensions is achieved by a data-driven dimensionality reduction strategy based on hierarchical functional decompositions. This facilitates a graph-theoretic representation of the active interactions in the data that can be used to guide the design of custom auto-correlation kernels. Such kernels can accurately explore spatial cross-correlations, while respecting the structure of interactions in the data. This observation is key as it enables the decomposition of the global high-dimensional machine learning problem into a series of independent low-dimensional optimization problems that can be efficiently formulated and solved in the frequency domain using $\mathcal{O}(N)$ parallel algorithms. Essentially, this opens the path to overcoming the computational deadlocks introduced by massive data-sets, which are often a necessity for statistical learning in high dimensional input spaces.

We have considered three benchmark problems that demonstrate the key features and capabilities of the proposed methodology. First, we have considered a stochastic flow problem, in which water flow through a borehole is described by two physical models of different fidelity taking values over an 8-dimensional input space. Stochas-

ticity is introduced by a Gaussian noise term depending on four random variables, leading to a response that is a 12-dimensional random field. Using a 4-level recursive co-kriging scheme we were able to accurately reconstruct the response surface of the mean field by integrating out the random noise effects using two approximation methods of different fidelity in probability space. This multi-fidelity information fusion approach resulted to an accurate predictive scheme that was trained on very few samples of the highest fidelity model, supplemented by more samples of inaccurate, lower fidelity realizations. Second, we presented results on the multi-fidelity regression of the Trid function in 100 input dimensions. In this case, a robust 3-level co-kriging predictor was built on few training samples coming from a high-fidelity model returning the exact function values, and an additional set of low-fidelity training data points originating from perturbations of the exact solution. The last test case demonstrates an extreme scenario of training a kriging scheme on 10^5 data-points in 1,000 input dimensions. This massive data-set was generated by randomly sampling the tensor product Sobol function in 1,000 dimensions, and the resulting additive kriging predictor was able to accurately represented the exact solution at several test locations. The key characteristic manifested by all benchmark cases is the high accuracy and computational tractability of the proposed framework both in terms of CPU cost as well as in terms of memory requirements.

Although our numerical results indicate that the proposed methodology is robust with respect to misspecification of the RS-HDMR expansion, a potential limiting factor affecting the accuracy of the predictions arises in cases where the RS-HDMR representation may not accurately detect the active interactions in the data. Such cases occur in very high-dimensional problems ($d > 100$) when only a limited number of observations is available. Performing meaningful inference in these data-poor environments becomes a daunting task, and is currently a subject of active research.

Although our current presentation mainly employed simplistic sampling strategies, such as uniform and Latin Hypercube sampling, we believe that the design of more sophisticated adaptive sampling techniques will be key to overcoming the aforementioned limitation.

CHAPTER FIVE

Modeling Continuum Blood Flow and Biomechanics

5.1 Motivation

The emerging multidisciplinary approach to modeling complex biological systems has had a big impact over the recent years in the development of numerical methods and algorithms for studying such systemic processes that span a large spectrum of spatial and temporal scales. Our particular area of interest here is arterial blood flow modeling, where fluid solvers have become a standard research tool for analyzing and predicting the underlying dynamics in scales ranging from the continuum (larger systemic arteries) to the atomistic level (small arterioles and veins). In this work, the goal is to study and analyze the potentiality and implementation aspects of using one-dimensional models as representations of the downstream vasculature, coupled as outflow boundary conditions to three-dimensional domains. The need of considering such reduced order models arises from the problem size in conjunction with the barrier imposed by the available computational resources. To this end, resolving the full problem in three dimensions involves $O(10^9)$ degrees of freedom and more than $10Tb$ of computer memory. This conflict can be greatly alleviated if one-dimensional or lumped parameter models are employed for closing the problem.

This chapter takes aim at three directions. First, we briefly present the formulation of one-dimensional (1D) models. Second, we present an overview of a spectral/*hp* element scheme for resolving three-dimensional (3D) continuum blood flow in fixed, as well as flexible arteries. Lastly, we propose and test an algorithm for the weak coupling of 3D-1D blood flow models. This exposition sets the ground for the contributions presented in Chapters [6](#), [7](#), and [8](#).

5.2 Nonlinear 1D blood flow models

We consider viscous incompressible 1D flow in a compliant tube. The flow dynamics is governed by a nonlinear hyperbolic system of partial differential equations that can be directly derived from the Navier-Stokes equations under the assumptions of axial symmetry, dominance of the axial velocity component, radial displacements of the arterial wall, and constant internal pressure on each cross section (see [Formaggia et al. \(2010\)](#)).

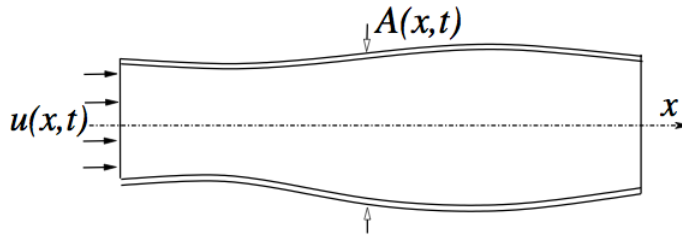


Figure 5.1: *Nonlinear 1D blood flow*: Flow in a 1D compliant artery (from [Sherwin et al. \(2003\)](#)).

The conservation of mass and momentum can be formulated in space-time (A, U) variables as (see [Sherwin et al. \(2003\)](#) for a detailed derivation) :

$$\begin{cases} \frac{\partial A}{\partial t} + \frac{\partial(AU)}{\partial x} = 0 \\ \frac{\partial U}{\partial t} + U \frac{\partial U}{\partial x} = -\frac{1}{\rho} \frac{\partial p}{\partial x} + K_r \frac{U}{\rho A}, \end{cases} \quad (5.1)$$

where x is the axial coordinate across the vessel's length, t is time, $A(x, t)$ is the cross-sectional area of the lumen, $U(x, t)$ is average axial fluid velocity, $Q(x, t) = AU$ is the mass flux, $p(x, t)$ is the internal pressure averaged over the tube's cross-section, and K_r is a friction parameter that depends on the velocity profile chosen (see [Sherwin](#)

et al. (2003)). Here we use an axisymmetric fluid velocity $u(x, r, t)$ profile that satisfies the no-slip condition.

$$u(x, r, t) = U \frac{\zeta + 2}{\zeta} \left\{ 1 - \left(\frac{r}{R} \right)^\zeta \right\}, \quad (5.2)$$

with $R(x, t)$ being the lumen radius, ζ a constant, and r the radial coordinate. Following Formaggia et al. (2010), $\zeta = 9$ gives a good fit to experimental blood flow data and $\zeta = 2$ returns the parabolic flow profile. Moreover, $U = \frac{1}{R^2} \int_0^R 2rudr$, and the friction parameter can be expressed as $K_r = \frac{2\mu}{U} \frac{A}{R} \left[\frac{\partial u}{\partial r} \right]_R = -22\mu\pi$ (see Sherwin et al. (2003)), with μ being the blood viscosity that is a function of the lumen radius and the blood hematocrit, based on the findings of Pries et al. (1992).

In order to close the system of Eq. 5.1, we need to provide a pressure area-relation. The simplest choice is to model the arterial wall as a purely elastic material. Then, the Laplace tube law yields (see Sherwin et al. (2003))

$$p = p_{ext} + \frac{\sqrt{\pi} E h}{(1 - \nu^2) \sqrt{A_0} \sqrt{A}} (\sqrt{A} - \sqrt{A_0}) \quad (5.3)$$

Here, p_{ext} is the external pressure on the arterial wall, E is the Young modulus of the wall, h is the wall thickness, and ν is the Poisson ratio.

5.2.1 Accounting for arterial viscoelasticity

We can further extend the capabilities of this formulation to capture more realistic viscoelastic arterial wall behaviors. In particular, while most 1-D models treat such interactions as purely elastic, the arterial walls are anisotropic and heterogeneous with complex nonlinear stress-strain relationships that give rise to phenomena such

as creep, stress relaxation and hysteresis (see [Fung \(1990\)](#)). Even though the elastic response is dominant for most vessels, the viscoelastic properties of the wall can have a significant impact on the flow, especially in distal locations (see [Reymond et al. \(2011\)](#)). The Voigt model is the simplest viscoelastic model that features creep and hysteresis. It depends only on the viscoelastic modulus of the wall, γ ; which makes it appealing for calibration to patient-specific data. However, if a more accurate representation of viscoelasticity is desired, a more involved modeling approach needs to be considered. The Kelvin model is a three-parameter linear viscoelastic model that serves exactly this purpose. It provides a discrete relaxation spectrum and enables us to resolve all the aforementioned viscoelastic effects. The only drawback is the inherent difficulty in estimating the required parameters from in-vivo data. However, this can be partially circumvented by employing a continuous relaxation spectrum.

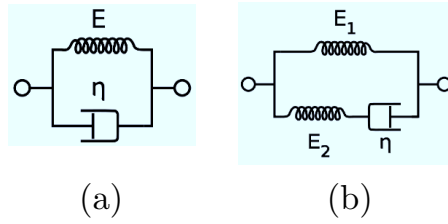


Figure 5.2: *Nonlinear 1D blood flow*: viscoelastic models considered. *Left*: Voigt. *Right*: Kelvin.

5.2.1.1 The Voigt arterial wall model

The stress-strain relation for a Voigt-type material is given by:

$$\sigma = E\epsilon + \eta \frac{d\epsilon}{dt} \quad (5.4)$$

where

$$\sigma = \frac{r(p - p_{ext})}{h} \quad \text{and} \quad \epsilon = \frac{1}{1 - \nu^2} \frac{r - r_0}{r_0} \quad (5.5)$$

are the stress and the strain respectively, while η is the viscoelastic modulus of the wall.

Substituting Eq. 5.5 into 5.4 and using the conservation of mass, one arrives at the Voigt pressure-area relation:

$$p = p_{ext} + \beta(\sqrt{A} - \sqrt{A_0}) - \frac{\Gamma}{\sqrt{A}} \frac{\partial(AU)}{\partial x} \quad (5.6)$$

with

$$\beta = \frac{\sqrt{\pi} E h}{(1 - \nu^2) A_0}, \quad \Gamma = \frac{2}{3} \frac{\sqrt{\pi} \eta h}{A_0}$$

Note that the inclusion of viscoelasticity adds a dissipative term to our system. In this case the system of Eqs. 5.1, 5.6 is no longer hyperbolic. However, the elastic contribution is still dominant thus allowing us to consider exactly the same numerical scheme used for the purely elastic case. Details of this implementation are provided in Sec. 5.2.2.

5.2.1.2 Fung's QLV and the Kelvin model

Here we present an overview of the theory of Quasi Linear viscoelasticity (QLV). The reader is referred to Fung (1990) for a detailed exposition on biomechanics. Modeling the response of soft tissue is often characterized by a constant hysteresis over a wider frequency range. Based on this approach we have a stress-strain relation of the form:

$$\sigma(x, t) = \int_0^\infty G(t - \gamma) \frac{\partial \epsilon^e(x, \gamma)}{\partial \gamma} d\gamma \quad (5.7)$$

where $G(t)$ is the stress relaxation function and $\epsilon^e(x, t)$ is the static elastic response of the tissue.

The presence of the convolution integral in Eq. 5.7 makes the stress depended upon the strain time history. It is the choice of the stress relaxation function that ultimately determines the model. In particular, we have:

- Choosing: $G(t - \gamma) = \eta \delta(t - \gamma)$, we get the Voigt model described in the previous section (δ is the Dirac delta function).

- Choosing:

$$G(t - \gamma) = 1 + \frac{\tau_\epsilon - \tau_\sigma}{\tau_\sigma} e^{-\frac{(t-\gamma)}{\tau_\sigma}}, \quad (5.8)$$

where

$$\tau_\sigma = \frac{\eta}{E_2} \quad \text{and} \quad \tau_\epsilon = \eta \frac{E_1 + E_2}{E_1 E_2} \quad (5.9)$$

yields the Kelvin model. Note that τ_ϵ is the relaxation time of constant strain and τ_σ is the relaxation time of constant stress. The Kelvin model is the simplest viscoelastic model that can capture effects of creep, stress relaxation and storage of strain energy at equilibrium. It is a linear model that employs only two discrete relaxation times.

- Choosing:

$$G(t - \gamma) = \frac{1 + c \int_{\tau_1}^{\tau_2} \frac{1}{\tau} e^{-\frac{(t-\gamma)}{\tau}} d\tau}{1 + c \int_{\tau_1}^{\tau_2} \frac{1}{\tau} d\tau}$$

yields a quasi-linear viscoelastic model with a continuous relaxation spectrum within the finite time interval $[\tau_1, \tau_2]$.

Using Eq. 5.5 and the Kelvin stress relaxation function of Eq. 5.8, we can integrate Eq. 5.7 by parts to arrive at the Kelvin pressure-area relation:

$$p(x, t) = p_{ext} + \frac{\tau_\epsilon}{\tau_\sigma} p^e(x, t) + \frac{\tau_\sigma - \tau_\epsilon}{\tau_\sigma^2} \int_0^t e^{-\frac{(t-\gamma)}{\tau_\sigma}} p^e(\gamma) d\gamma, \quad (5.10)$$

where

$$p^e(x, t) = \beta(\sqrt{A(x, t)} - \sqrt{A_0})$$

is the static elastic response, and $\beta = \frac{\sqrt{\pi} E h}{(1 - \nu^2) A_0}$.

As in the Voigt model, the pressure-area relation of Eq. 5.10 introduces dissipation to our system, thus violating hyperbolicity. Again, the elastic response is an order of magnitude greater than the viscoelastic, hence the numerical method will be a modification of the one used for the elastic case.

The following table summarizes the features of each arterial wall model presented above:

Model	Creep	Stress Relaxation	Hysteresis	Rel. Spectrum
Elastic	-	-	-	-
Voigt	✓	-	✓	-
Kelvin	✓	✓	✓	$\{\tau_\epsilon\}, \{\tau_\sigma\}$
General QLV	✓	✓	✓	$[\tau_1, \tau_2]$

5.2.2 The Discontinuous Galerkin scheme

The system of equations in Eq. 5.1 can be recast in a conservative form as (see Sherwin et al. (2003)):

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} = \mathbf{S}(\mathbf{U}), \quad \mathbf{U} = \begin{bmatrix} A \\ U \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} AU \\ \frac{U^2}{2} + \frac{p^e}{\rho} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{p^v}{\rho} \end{bmatrix}, \quad \mathbf{S}(\mathbf{U}) = \begin{bmatrix} 0 \\ -22\mu\pi \frac{U}{\rho A} \end{bmatrix}, \quad (5.11)$$

where $p^v(x, t)$ is the viscoelastic component of the total pressure. Here, in order to analyze the system, we consider the case of zero forcing a purely elastic pressure-area relation (Eq. 5.11 with $p^v(x, t) = 0$). Then, the system of Eq. 5.11 forms a non-linear hyperbolic system, which in conservative form reads as

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{H}(\mathbf{U}) \frac{\partial \mathbf{U}}{\partial x} = 0, \quad (5.12)$$

where

$$\mathbf{U} = \begin{bmatrix} A \\ u \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} u & A \\ \frac{1}{\rho D A} & u \end{bmatrix}, \quad (5.13)$$

and the arterial distensibility D is defined as

$$D = \frac{1}{A} \frac{dA}{dp} = \frac{2}{\beta \sqrt{A}}. \quad (5.14)$$

Under the assumption that $A > 0$, the matrix $\mathbf{H}$ has two real eigenvalues:

$$\lambda_{1,2}(\mathbf{H}) = u \pm c, \quad c = \frac{1}{\sqrt{\rho D}}. \quad (5.15)$$

Then the matrix of left eigenvectors of $\mathbf{H}$ is:

$$\mathbf{L} = \begin{bmatrix} \frac{c}{A} & 1 \\ -\frac{c}{A} & 1 \end{bmatrix}, \quad (5.16)$$

where

$$\mathbf{LH} = \mathbf{\Lambda L}, \quad \text{with,} \quad \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad \text{and} \quad \mathbf{H} = \mathbf{L}^{-1} \mathbf{\Lambda L} \quad (5.17)$$

Hence we obtain:

$$\mathbf{L} \frac{\partial \mathbf{U}}{\partial t} + \mathbf{\Lambda L} \frac{\partial \mathbf{U}}{\partial x} = 0, \quad (5.18)$$

which can be decoupled by changing variables $\frac{\partial \mathbf{W}}{\partial \mathbf{U}} = \mathbf{L}$, $\mathbf{W} = [W_1, W_2]^T$:

$$\begin{cases} \frac{\partial W_1}{\partial t} + \lambda_1 \frac{\partial W_1}{\partial x} = 0 \\ \frac{\partial W_2}{\partial t} + \lambda_2 \frac{\partial W_2}{\partial x} = 0 \end{cases} \quad (5.19)$$

leading to an analytical expression for the forward and backward Riemann invariants, respectively

$$W_{1,2} = U \pm 4 \sqrt{\frac{\beta}{2\rho}} \left(A^{1/4} - A_0^{1/4} \right) \quad (5.20)$$

The computational domain Ω consists of arterial segments, which can be divided in N_{el} elemental non-overlapping regions $\Omega_e = (x_e^L, x_e^R)$, such that $x_e^R = x_{e+1}^L$ for $e = 1, \dots, N_{el}$. The discontinuous Galerkin formulation requires, for each element $e = 1, \dots, N_{el}$, the resolution of the system:

$$J_e \frac{\partial \hat{U}_{i,e}^p}{\partial t} = -J_e \int_{\Omega_e} L_p \frac{\partial F_i}{\partial x} d\xi - L_p [F_i^u - F_i]_{x_e^L}^{x_e^R} + J_e \int_{\Omega_e} L_p S_i d\xi, \quad p = 0, \dots, P, \quad i = 1, 2, \quad (5.21)$$

where $\mathbf{U}$ has been discretized by $\mathbf{U}^\delta$ written in terms of orthonormal Legendre polynomials $L_p(x)$:

$$\mathbf{U}^\delta|_{\Omega_e} = \sum_{p=0}^P L_p \hat{\mathbf{U}}_e^p$$

Each element is mapped onto a reference element $\Omega_{st} = \{-1 \leq \xi \leq 1\}$ called the "standard element" and J_e is the Jacobian of the corresponding affine mapping:

$$J_e = \frac{1}{2}(x_e^R - x_e^L), \quad x_e(\xi) = x_e^L \frac{(1 - \xi)}{2} + x_e^R \frac{(1 + \xi)}{2}$$

$\mathbf{F}^u$ is the upwinded flux that propagates information between the elemental regions and the bifurcations of the system. At the inlet and outlet boundary elements, the fluxes are upwinded by means of the boundary conditions. The hyperbolicity of the system requires one boundary condition at each terminal end.

An Adams-Bashforth scheme is used for the time integration:

$$\frac{\partial \hat{U}_{i,e}^p}{\partial t} = f(\hat{U}_{i,e}^p)$$

with,

$$\left(\hat{U}_{i,e}^p\right) = - \int_{\Omega_e} L_p \frac{\partial F_i}{\partial x} d\xi - \frac{1}{J_e} L_p [F_i^u - F_i]_{x_e^L}^{x_e^R} + \int_{\Omega_e} L_p S_i d\xi,$$

then,

$$\left(\hat{U}_{i,e}^p\right)^{n+1} = \left(\hat{U}_{i,e}^p\right)^n + \frac{3\Delta t}{2} f\left(\left(\hat{U}_{i,e}^p\right)^n\right) - \frac{\Delta t}{2} f\left(\left(\hat{U}_{i,e}^p\right)^{n-1}\right)$$

In the case of a viscoelastic wall model, the flux $\mathbf{F}$ has to be separated in an elastic part $\mathbf{F}_e$ and a viscoelastic part so that: $\mathbf{F} = \mathbf{F}_e + \mathbf{F}_v$, where:

$$\mathbf{F}_e = \begin{bmatrix} uA \\ \frac{u^2}{2} + \frac{p^e}{\rho} \end{bmatrix}, \quad \mathbf{F}_v = \begin{bmatrix} 0 \\ \frac{p^v}{\rho} \end{bmatrix}$$

with $p^e = p_{ext} + \beta(\sqrt{A} - \sqrt{A_0})$ for the elastic and Voigt cases, while $p^e = p_{ext} + \frac{\tau_\epsilon}{\tau_\sigma} \beta(\sqrt{A} - \sqrt{A_0})$ for the Kelvin model.

In particular for the Voigt model we have

$$\mathbf{F}_v^{\text{Voigt}} = \begin{bmatrix} 0 \\ -\frac{\Gamma}{\sqrt{A}} \frac{\partial(AU)}{\partial x} \end{bmatrix}$$

Similarly, for the Kelvin model we have:

$$\mathbf{F}_v^{\text{Kelvin}} = \begin{bmatrix} 0 \\ \frac{\tau_\sigma - \tau_\epsilon}{\tau_\sigma^2} \int_0^t e^{-\frac{(t-\gamma)}{\tau_\sigma}} p^e(\gamma) d\gamma \end{bmatrix}$$

Normally, the evaluation of the convolution integral in the Kelvin pressure-area

relation requires a number of operations that scale quadratically with the number of considered history points. Following the approach of [Steele et al. \(2011\)](#), we can explore the properties of the exponential term, we can derive a simple update rule that is in-dependent of the number of history points considered: simple update rule:

$$I_{t+\Delta t} = \int_0^{t+\Delta t} e^{-\frac{(t+\Delta t-\gamma)}{\tau_\sigma}} p^e(\gamma) d\gamma = e^{-\frac{\Delta t}{\tau_\sigma}} I_t + \int_t^{t+\Delta t} e^{-\frac{(t+\Delta t-\gamma)}{\tau_\sigma}} p^e(\gamma) d\gamma$$

where the last term contributes for the time interval $[t, t + \Delta t]$ and is computed using the trapezoidal rule. This effectively allows us to consider the full history of the arterial wall response without introducing any computational bottleneck.

5.2.3 Boundary Conditions

At bifurcation and junction points of the network we require continuity of the Riemann invariants, conservation of mass and continuity of total pressure (see [Sherwin et al. \(2003\)](#) for a detailed analysis).

Terminal vessels are typically coupled to 0-D lumped parameter models. The most suitable choice for one dimensional blood flow simulations is the RCR model, where we specify the total peripheral resistance downstream of the outlet R_t , and a corresponding compliance C (see [Alastruey et al. \(2008\)](#)):

$$C \frac{dp_c}{dt} = A^* U^* - \frac{p_c - p_{out}}{R_2}, \quad R_2 = R_t - R_1, \quad R_1 \doteq \frac{c_0 \rho}{A_0}, \quad c_0 = \sqrt{\frac{1}{2} \frac{\beta}{\rho}} A_0^{\frac{1}{4}} \quad (5.22)$$

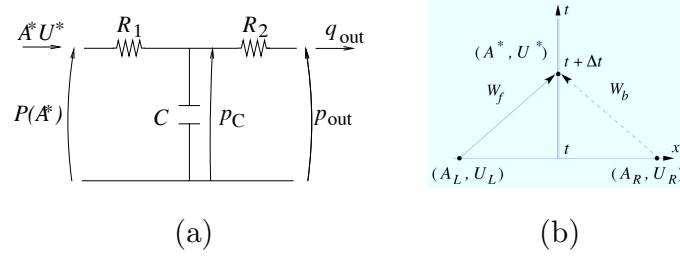


Figure 5.3: *Nonlinear 1D blood flow*: *Left*: Sketch of the RCR outflow model. *Right*: Sketch of the forward and backward characteristics at the 1D-0D coupling interface (figure taken from [Alastruey et al. \(2008\)](#)).

Following [Alastruey et al. \(2008\)](#), we solve the coupling as

$$\mathcal{F}(A^*) = \frac{\rho c_0}{A_0} [U_L + 4c(A_L)] A^* - 4 \frac{\rho c_0}{c} (A^*) A^* - \beta (\sqrt{A^*} - \sqrt{A_0}) + p_c = 0 \quad (5.23)$$

Finally, the outflow boundary condition is obtained by imposing $A_L = A_R$, and $U_R = 2U^* - U_L$.

5.3 Linear 1D blood flow models

The analysis can be further simplified if one adopts a linear formulation (see [Sherwin et al. \(2003\)](#), [Alastruey et al. \(2008\)](#)). This can be achieved by expressing System 5.11 in terms of the (A, p, Q) variables, with $Q = AU$, and linearizing them about the reference state $(A_0, 0, 0)$, with β and A_0 held constant along x . This yields

$$C_{1D} \frac{\partial \tilde{p}}{\partial t} + \frac{\partial \tilde{q}}{\partial x} = 0, \quad (5.24)$$

$$L_{1D} \frac{\partial \tilde{q}}{\partial t} + \frac{\partial \tilde{p}}{\partial x} = -R_{1D} \tilde{q}, \quad (5.25)$$

$$\tilde{p} = \frac{\tilde{a}}{C_{1D}}, \quad (5.26)$$

where $\tilde{a}$, $\tilde{p}$ and $\tilde{q}$ are the perturbation variables for area, pressure, and volume flux, respectively, i.e., $(A, p, Q) = (A + \tilde{a}, \tilde{p}, \tilde{q})$, and

$$R_{1D} = \frac{2(\zeta + 2)\pi\mu}{A_0^2}, \quad L_{1D} = \frac{\rho}{A_0}, \quad C_{1D} = \frac{A_0}{\rho c_0^2}, \quad (5.27)$$

are the viscous resistance to flow, blood inertia and wall compliance, respectively per unit length of vessel (see [Alastruey et al. \(2008\)](#)). For a purely elastic arterial wall behavior it follows that

$$\frac{\partial \tilde{p}}{\partial \tilde{a}} = \left. \frac{\partial p}{\partial A} \right|_{A=A_0} = \frac{1}{C_{1D}} \quad (5.28)$$

Then, the corresponding linear forward and backward Riemann invariant can analytically be expressed

$$W_{1,2} = \tilde{q} \pm \frac{\tilde{p}}{Z_0}, \quad Z_0 = \frac{\rho c_0}{A_0}, \quad (5.29)$$

where Z_0 is the characteristic impedance of the vessel.

5.4 3D blood flow models

5.4.1 Flow in rigid vessels

Here, we provide an overview of the sub-iterative scheme for solving the incompressible Navier-Stokes equations put forth by [Baek and Karniadakis \(2011\)](#). To this end, we consider 3D unsteady incompressible flow in a rigid domain Ω , described by the Navier-Stokes equations subject to the incompressibility constraint:

$$\begin{cases} \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot (\nabla \mathbf{v}) = -\nabla p + \nu \nabla^2 \mathbf{v} + \mathbf{f} \\ \nabla \cdot \mathbf{v} = 0 \end{cases} \quad (5.30)$$

where $\mathbf{v}$ is the velocity vector, p is the pressure, t is time, ν is the kinematic viscosity of the fluid, and $\mathbf{f}$ are the external body forces. The system is discretized in space using the spectral/ hp element method (SEM) (see [Karniadakis and Sherwin \(2013\)](#)), according to which, the computational domain Ω is decomposed into a set of polymorphic non-overlapping elements $\Omega^{e_i} \subset \Omega$, $i = 1, \dots, Nel$. Within each element, the solution is approximated as a linear combination of hierarchical, mixed-order, semi-orthogonal Jacobi polynomial expansions (see [Karniadakis and Sherwin \(2013\)](#)). This hierarchical structure consists of separated vertex (linear term) $\Phi_k(\mathbf{x})$, edge $\Psi_k(\mathbf{x})$, face $\Theta_k(\mathbf{x})$ and interior (or bubble) modes $\Lambda_k(\mathbf{x})$. According to this decomposition, the polynomial representation of a field $\mathbf{v}(t, \mathbf{x})$ at any point $\mathbf{x}_j$ is given by the linear combination of the basis functions multiplied by the corresponding modal amplitudes:

$$\mathbf{v}(t, \mathbf{x}_j) = \sum_{k=1}^{Nv} \hat{\mathbf{v}}_k^V(t) \Phi_k(\mathbf{x}_j) + \sum_{k=1}^{Ne} \hat{\mathbf{v}}_k^E(t) \Psi_k(\mathbf{x}_j) + \sum_{k=1}^{Nf} \hat{\mathbf{v}}_k^F(t) \Theta_k(\mathbf{x}_j) + \sum_{k=1}^{Ni} \hat{\mathbf{v}}_k^I(t) \Lambda_k(\mathbf{x}_j) \quad (5.31)$$

The numerical solution of the above system is based on decoupling the velocity and pressure by applying a high-order time-splitting scheme (see [Karniadakis and Sherwin \(2013\)](#)). We solve for an intermediate velocity field $\mathbf{v}^*$ in physical space (Equation 5.32) before a Galerkin projection is applied to obtain the weak formulation for the pressure and velocity variables (Equations 5.33, 5.34), which are computed by solving a Poisson and a Helmholtz problem, respectively. At time step n and sub-iteration step k we solve Equations 5.33, 5.34 for p_k^n and $\mathbf{v}_k^n$, respectively, using previous time-step solutions $\mathbf{v}^{n-q}$, $q = 1, \dots, J_e$, where J_e denotes the order of the stiffly stable time integration scheme used.

$$\mathbf{v}^* = \sum_{q=0}^{J_e-1} \alpha_q \mathbf{v}^{n-q} - \Delta t (\mathbf{N}_{k-1}^n + \mathbf{f}) \quad (5.32)$$

$$\mathbf{L} \hat{p}_k^n = -\frac{1}{\Delta t} (\nabla \cdot \mathbf{v}^*, \phi) + \left(\frac{\partial p_k^n}{\partial n}, \phi \right) - \mathbf{L}^D \hat{p}^D \quad (5.33)$$

$$\mathbf{H} \hat{\mathbf{v}}_k^n = \frac{1}{\gamma_0} (\mathbf{v}^* - \Delta t \nabla p_k^n, \phi) + \frac{\Delta t \nu}{\gamma_0} \left(\frac{\partial \mathbf{v}_k^n}{\partial n}, \phi \right) - \mathbf{H}^D \hat{\mathbf{v}}^{n+1,D} \quad (5.34)$$

where we define

$$\mathbf{N}_{k-1}^n = \mathbf{v}_{k-1}^n \cdot \nabla \mathbf{v}_{k-1}^n \quad (5.35)$$

$$\mathbf{V}_{k-1}^n = \nu \nabla \times (\nabla \times \mathbf{v}_{k-1}^n) \quad (5.36)$$

Also, γ_0 , α_k are the coefficients of a backward differentiation formula, β_k are the coefficients of the stiffly stable time integration scheme, $\mathbf{H} = \mathbf{M} - \frac{\Delta t \nu}{\gamma_0}$, and $\mathbf{M}$ and $\mathbf{L}$ are the mass and stiffness matrices, respectively. Moreover, $\hat{\mathbf{v}}$ and $\hat{p}$ are the unknown modal amplitudes of the velocity and pressure variables, while variables with a D superscript are considered to be known. In particular, the last term in Equations 5.33, 5.34 is due to lifting a known solution for imposing Dirichlet boundary conditions, and the corresponding operators $\mathbf{H}^D$ and $\mathbf{L}^D$ need to be constructed only for these boundaries (see Karniadakis and Sherwin (2013)). The system is closed with a Neumann $\partial \mathbf{v} / \partial n = 0$ boundary condition for the velocity at the outlets, and a consistent Neumann pressure boundary condition at the boundaries with a prescribed Dirichlet condition for the velocity field (see Karniadakis and Sherwin (2013)):

$$\frac{\partial p_k^n}{\partial n} = - \left[\frac{\gamma_0 \mathbf{v}_k^n - \sum_{q=0}^{Je-1} \alpha_k \mathbf{v}^{n-q}}{\Delta t} + \sum_{k=0}^{Je-1} [\beta_k (\mathbf{N}_{k-1}^n + \mathbf{V}_{k-1}^n)]^{n-k} \right] \cdot \mathbf{n} \quad (5.37)$$

At the first sub-iteration step $k = 1$, the solution from previous sub-iterations is not available and $\mathbf{N}_0^n, \mathbf{V}_0^n$ are extrapolated in time as follows:

$$\mathbf{N}_0^n = \sum_{q=1}^{J_e} \beta_q [\mathbf{v} \cdot \nabla \mathbf{v}]^{n-q} \quad (5.38)$$

$$\mathbf{V}_0^n = \nu \sum_{q=1}^{J_e} \beta_q [\nabla \times (\nabla \times \mathbf{v})]^{n-q} \quad (5.39)$$

After computing $\hat{\mathbf{v}}_k^n$ and $\hat{p}_k^n$ we can accelerate the convergence of the iterative procedure by applying the Aitken under-relaxation scheme to the computed velocity field:

$$\tilde{\mathbf{v}}_k^n = \lambda_k \tilde{\mathbf{v}}_{k-1}^n + (1 - \lambda_k) \mathbf{v}_k^n \quad (5.40)$$

where $\lambda_k \in [0, 1]$ is computed using the classical Aitken rule:

$$\lambda_k = \lambda_{k-1} + (\lambda_{k-1} - 1) \frac{(\mathbf{Q}_{k-1} - \mathbf{Q}_k) \cdot \mathbf{Q}_k}{\|\mathbf{Q}_{k-1} - \mathbf{Q}_k\|^2}, \quad \mathbf{Q}_k = \tilde{\mathbf{v}}_{k-1}^n - \mathbf{v}_k^n \quad (5.41)$$

Finally, convergence is checked by requiring $\|\mathbf{v}_k^n - \mathbf{v}_{k-1}^n\|$ and $\|p_k^n - p_{k-1}^n\|$ to be less than a pre-specified tolerance.

For a complete analysis of this scheme the reader is referred to [Karniadakis and Sherwin \(2013\)](#), [Baek and Karniadakis \(2011\)](#). Results in the context of arterial blood flow simulations can be found in [Grinberg et al. \(2011a\)](#), while parallel performance has been documented in [Grinberg et al. \(2011b\)](#).

5.4.1.1 Treatment of boundary conditions

Given the Fourier representation of the volumetric flow-rate (VFR) measurements at the ascending aorta, the inflow is driven by a time-dependent velocity boundary condition based on the Womersley profile. This model takes into account the pulsatile nature of the flow and it corresponds to the exact solution of pulsatile flow in a straight rigid pipe.

At the outlets we employ a lumped parameter RC model that results to a Dirichlet pressure boundary condition which is determined by the VFR $Q(t)$, the resistance R and the capacitance C (see [Grinberg and Karniadakis \(2008\)](#)):

$$p(t) + RC \frac{dp(t)}{dt} = Q(t)R \quad (5.42)$$

As reported by [Grinberg and Karniadakis \(2008\)](#), the advantage of this boundary

condition is the attenuation of high frequency oscillations. Our approach in choosing the R and C parameters follows the analysis presented by [Grinberg and Karniadakis \(2008\)](#), according to which we set an arbitrary value for the resistance R_j at the j^{th} outlet and we calculate the resistances at all other outlets as:

$$R_i = R_j \frac{\bar{Q}_j}{\bar{Q}_i} \quad (5.43)$$

by utilizing the given patient-specific mean VFR measurements at these locations.

5.4.2 Flow-structure interactions

We consider the solution of the 3D Navier-Stokes equations taking into account arterial compliance. Given proper initial conditions, the system of equations describing the fluid flow is derived in the arbitrary Lagrangian-Eulerian (ALE) reference frame as (see [Baek and Karniadakis \(2012\)](#), [Yu et al. \(2013\)](#))

$$\begin{cases} \left. \frac{\partial^\star \mathbf{v}}{\partial t} \right|_{\mathcal{M}} + (\mathbf{v} - \mathbf{w}) \cdot (\nabla \mathbf{v}) = -\nabla p + \nu \nabla^2 \mathbf{v} + \mathbf{f} \\ \nabla \cdot \mathbf{v} = 0, \text{ in } \Omega(t), \quad \mathbf{u} = \mathbf{v} = \mathbf{w}, \mathbf{x} \in \partial\Omega_s, \end{cases} \quad (5.44)$$

where $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are the structure, fluid and mesh velocities, respectively, and $\partial\Omega_s$ denoted the fluid-solid interface. The $(\star)$ symbol implies that the derivative is taken on the mesh frame $\mathcal{M}$. At the interface between fluid subdomain and solid subdomain, Dirichlet and Neumann type boundary conditions are imposed for the velocity and pressure, respectively.

To model the response of the arterial wall, consider a structure initially occupying domain Ω_0 being deformed to a new configuration Ω at time t . Let $\mathbf{X}$ and $\mathbf{x}(\mathbf{X}, t)$

denote the coordinate in the undeformed configuration and the displacement at the location $\mathbf{X}$ and at time t . Then such deformation can be described as

$$\begin{cases} \rho \frac{\partial^2 \mathbf{x}}{\partial t^2} = \nabla \cdot \boldsymbol{\tau} + \rho \mathbf{b}, & \mathbf{x} \in \Omega \\ \mathbf{t} = \boldsymbol{\tau} \cdot \mathbf{n}, & \mathbf{x} \in \partial\Omega, \end{cases} \quad (5.45)$$

where $\boldsymbol{\tau}$, $\mathbf{b}$, $\mathbf{t}$ and ρ are the stress tensor, external body force, external traction force, and structural mass density, respectively, and $\mathbf{x}(\mathbf{X}, t)$ is the unknown displacement. Under the assumption that deformation is small and material is linear elastic, the stress tensor $\boldsymbol{\tau} = C\boldsymbol{\epsilon}$, where C is the material matrix with two material constants, Young's modulus (E) and Poisson constant (ν) and $\boldsymbol{\epsilon} = 1/2(\nabla \mathbf{x} + \nabla \mathbf{x}^T)$, where T indicates the transpose of a tensor or matrix. For extensions to hyper-elastic and viscoelastic constitutive laws the reader is referred to [Yu et al. \(2012\)](#).

The fluid and solid domains are coupled through kinematic and dynamics boundary condition at the interface

$$\begin{cases} \mathbf{v}_f \cdot \mathbf{n}_f = -\mathbf{u}_s \cdot \mathbf{n}_s, \\ \boldsymbol{\sigma}_f \cdot \mathbf{n}_f = -\boldsymbol{\sigma}_s \cdot \mathbf{n}_s \end{cases} \quad (5.46)$$

The above equations are discretized using the spectral element method put forth by [Baek and Karniadakis \(2012\)](#), [Yu et al. \(2013\)](#). As described, the fluid solver employs the ALE formulation and updates the mesh at every time step. The solid solver employs an elastodynamic equation with traction boundary condition at the fluid interface. Both solvers are based on third-order accurate time-stepping schemes. Stability and convergence of the solution are enhanced by performing sub-iterations in conjunction with Aitken relaxation. The relaxation procedure is crucial as it leads to stable coupling even for small mass ratios – a common problem in FSI simulations,

recently addressed by [Yu et al. \(2013\)](#).

5.5 Coupling of the 3D and 1D solvers

In this section we aim to explore the coupling of nonlinear 1D model to 3D flow simulations in rigid arteries. While reduced order models, such as 0D lumped parameter models and 1D linearized models have been used extensively in the past as outflow boundary conditions, here our ultimate goal is to model the downstream vasculature as a fractal network (see Chapter 7, and [Zamir \(1999\)](#)), in which the fully non-linear 1D problem is solved. In what follows, we present the main algorithm along with the sets of possible coupling conditions, and several test cases.

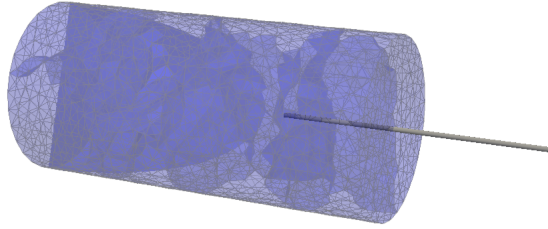


Figure 5.4: *3D-1D coupling*: Computational domain for 3D-1D overlapping coupling for a straight pipe.

We consider a rigid tube 3D domain Ω_{3D} with its outlet being coupled to a compliant 1D vessel Ω_{1D} (with or without an overlapping region). At each time-step, the hyperbolic 1D model requires one boundary condition at its inlet, while the parabolic 3D model requires two boundary conditions at its outlet. A detailed discussion on such coupling conditions, well-posedness and stability of the resulting system has been given by [Formaggia et al. \(2010\)](#), [Formaggia et al. \(2001\)](#), [Passerini et al. \(2009\)](#). Among the admissible choices presented therein, we consider the following cases:

- Case 1 (defective):

$$\begin{cases} Q_{1D}^n = Q_{3D}^n, \\ p_{3D}^{n+1} = p_{1D}^{n+1}, \\ \frac{\partial \mathbf{v}^{n+1}}{\partial n} = \frac{du_{1D}^{n+1}}{dx} \end{cases} \quad (5.47)$$

- Case 2 (defective):

$$\begin{cases} Q_{1D}^n = Q_{3D}^n, \\ p_{3D}^{n+1} = f(R_{1D}^{n+1}, C_{1D}^{n+1}), \\ \frac{\partial \mathbf{v}^{n+1}}{\partial n} = \frac{du_{1D}^{n+1}}{dx} \end{cases} \quad (5.48)$$

- Case 3:

$$\begin{cases} Q_{1D}^n = Q_{3D}^n, \\ \{-(p_{3D}^{n+1} + \frac{\rho}{2}|u_{3D}^n|^2)I - \mu(\nabla u_{3D}^n + \nabla u_{3D}^{nT})\} \cdot \mathbf{n} = \{p_{1D}^{n+1} + \frac{\rho}{2}|u_{1D}^{n+1}|^2\} \cdot \mathbf{n}, \\ \frac{\partial \mathbf{v}^{n+1}}{\partial n} = \frac{du_{1D}^{n+1}}{dx} \end{cases} \quad (5.49)$$

- Case 4:

$$\begin{cases} Q_{1D}^n = Q_{3D}^n, \\ W_2 = U_{3D}n - 2\sqrt{\frac{\rho}{2}}(\sqrt{p_{3D} + \beta\sqrt{A_0}} - \sqrt{\beta\sqrt{A_0}}), \\ \frac{\partial \mathbf{v}^{n+1}}{\partial n} = \frac{du_{1D}^{n+1}}{dx} \end{cases} \quad (5.50)$$

where $Q = Au$ is the flow rate, A the cross sectional area, u the velocity component in the axial direction, p the pressure, W_2 the incoming characteristic from the 1D domain, $R = \frac{p}{Q}$ the resistance and C the corresponding compliance. Note that suggested sets of coupling conditions only differ in the treatment of the pressure Dirichlet boundary condition applied at the outlet of the 3D domain.

In case 1, the 1D model is driven by the flow-rate Q_{3D} (imposed as an inflow

Dirichlet boundary condition), while it returns the pressure p_{1D} and the flux $\frac{du}{dx}$ to be imposed as Dirichlet and Neumann outflow boundary conditions at the 3D outlet, respectively.

In case 2, the idea is exactly the same as in case 1 but here we impose the Dirichlet boundary condition for the pressure at the 3D outlet using a low pass RC-filter to attenuate any spurious oscillations coming from the downstream 1D model. In comparison to case 1, this leads to faster convergence of the CG solver for the pressure in the 3D domain when time-extrapolation is used to accelerate the 3D solver (see [Grinberg and Karniadakis \(2011\)](#)), thus reducing the total computation time. Note that since the 1D domain can only provide a single averaged value for the pressure and the outflow flux (instead of point-wise data naturally required by the 3D domain), cases 1,2 give rise to a defective boundary condition problem.

In case 3, the traction at the 3D outlet is set to be equal to the total pressure at the 1D inlet. In the cases we study here, this formulation is identical to the conditions listed in case 1. However, when the non-linear terms in the Navier-Stokes are written in rotational form, it provides a non-defective outflow boundary condition as it can return point-wise data for the pressure (see [Formaggia et al. \(2001\)](#)).

In case 4, point-wise pressure data is recovered by exploiting the continuity of the forward and backward Riemann invariants at the coupling interface. However, it is only applicable for the case of coupling compliant 3D-1D coupling (see [Papadakis \(2009\)](#)).

In what follows we will confine our study to the rigid 3D - flexible 1D coupling, thus we will only employ, test and compare cases 1,2 of the coupling conditions presented above.

5.5.1 The coupling algorithm

The goal is to utilize the 1D model for providing the 3D domain with an accurate pressure boundary condition at time level $(n + 1)$. To this end, we need to integrate the 1D domain from t^n to t^{n+1} driven by the inflow imposed by Q_{3D}^n .

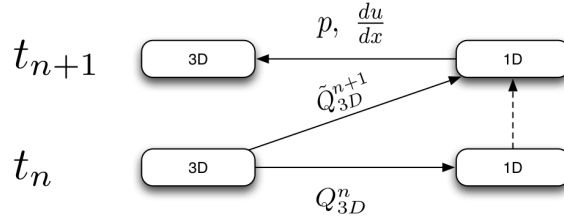


Figure 5.5: *3D-1D coupling*: Sketch of the 3D-1D weak coupling algorithm

Typically the explicit integration scheme of the hyperbolic 1D system requires a smaller time-step than the one used in the 3D counterpart. For more accurate results we can provide the 1D domain with a time-extrapolated flow-rate $\tilde{Q}_{3D}^{n+1}$ to be used for the intermediate 1D time-steps:

$$\tilde{Q}_{3D}^{n+1} = f(Q_{3D}^n, Q_{3D}^{n-1}, Q_{3D}^{n-2})$$

Hence, between t_n and t_{n+1} the flow-rate at the 1D inlet can be computed using interpolation between Q_{3D}^n and $\tilde{Q}_{3D}^{n+1}$ as shown in Fig.6:

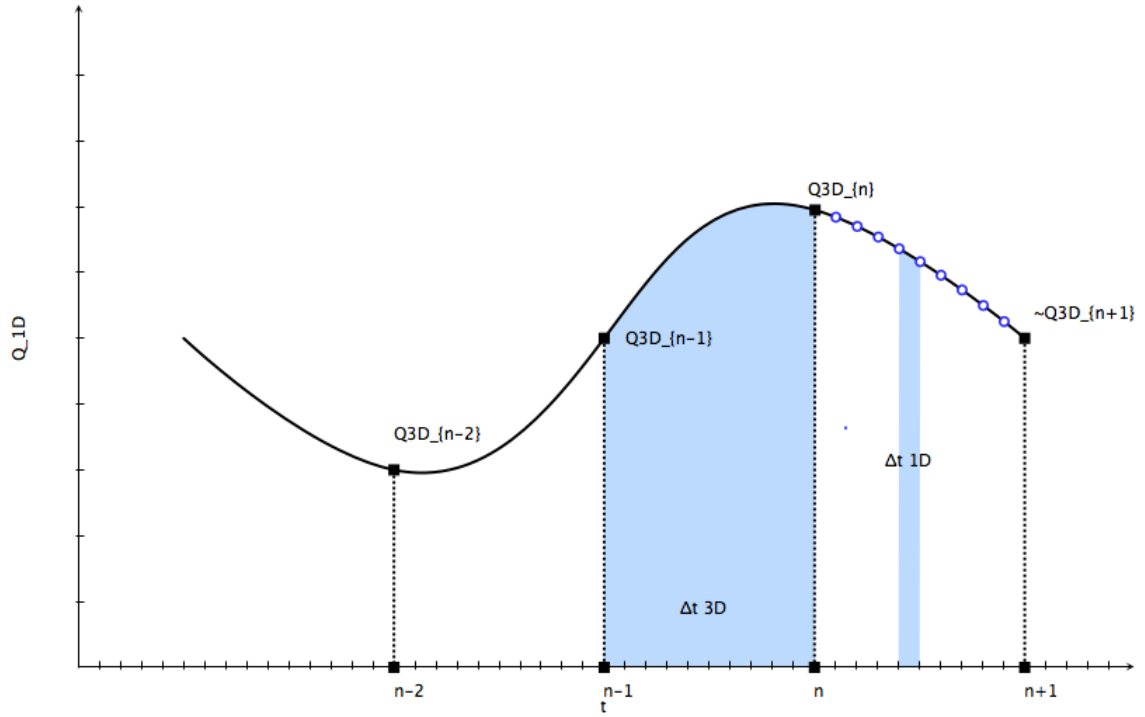


Figure 5.6: *3D-1D coupling*: Flow-rate interpolation for intermediate asynchronous 1D solver steps.

Finally, at t^{n+1} we can provide the 3D domain with the corresponding boundary conditions for the outflow pressure and flux (either directly or through a low-pass RC filter) and solve for the field variables $\mathbf{v}$ and p as described in Section 2.

We have implemented and tested this coupling using the spectral element solver NektarG for the 3D domain, and the nonlinear 1D Discontinuous Galerking solver described in Sec. 5.2 for the 1D domain. The implementation of the required communication between the 3D and 1D solvers is based on non-blocking calls of the Message Passing Interface library (MPI) which involve only the CPU ranks handling elements on each coupling interface. The overlap between communication and computation makes the additional computational cost of this implementation negligible as long as the size of the 1D network is relatively small (up to a few thousands of vessels). For 1D networks consisting of tenths of thousands vessel or even millions, the serial 1D

solver causes a computation bottleneck. In Chapter 7 of this thesis we will address this issue by introducing a parallel implementation of the 1D solver that allows near real-time solutions in fractal arterial networks with millions of arteries. Finally, we note that a strong coupling between the geometrically heterogeneous solvers can be achieved by the means of sub-iterations of the aforementioned partitioned algorithm.

5.5.2 Remarks

5.5.2.1 Overlapping vs non-overlapping coupling

All results on 3D-1D coupling that can be found in the literature have employed a non-overlapping domain decomposition strategy. Here, motivated by the work of Grinberg et al. (2011b), we introduce and analyze the potentiality of using an overlapping domain region between the 3D and 1D models. In the context of spectral methods, Karniadakis and Sherwin (2013) have shown that the time-splitting scheme used for the 3D model introduced a numerical boundary layer of thickness $\delta_t \approx (\nu \Delta t)^{J_e}$ that decays exponentially over the distance s away from the domain boundary (error $\propto e^{-s/\delta_t}$). As demonstrated by the results presented by Grinberg et al. (2011b), extracting the coupling quantities from the inner elements of the two communicating domains reduces the numerical boundary layer errors and enhances the stability of the coupling. As we will see in the Results section, this approach indeed returns more accurate results, but more importantly, overcomes the stability issues that the non-overlapping setup may suffer from.

5.5.2.2 Imposing the outflow flux $\frac{\partial \mathbf{v}}{\partial n}$

A standard way for imposing the Neumann outflow boundary condition in stand-alone 3D flow solvers is to assume $\frac{\partial \mathbf{v}}{\partial n} = 0$, where n is the normal direction to the outlet's cross section. This may not be the case for unsteady flow in geometries different than the straight tube. To demonstrate this we set up a simple simulation of Womersley flow in a pipe with 5° tapering along the axial direction and compared the results against the straight pipe case.

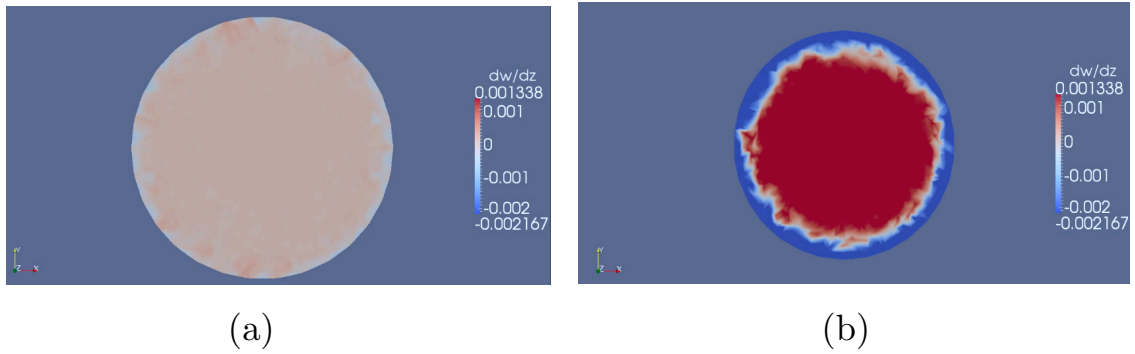


Figure 5.7: *3D-1D coupling*: Normal flux close to the outflow boundary of a straight pipe (*left*) and a converging pipe with 5 degrees tapering (*right*).

As we can see in 5.7 the normal flux close to the outflow boundary takes values away from zero for the diverging pipe case, indicating that imposing $\frac{\partial \mathbf{v}}{\partial n} = 0$ as a Neumann outflow boundary condition is not consistent.

When coupling a 3D domain to a 1D network, we can utilize the information coming from the downstream 1D domain in order to impose a more accurate Neumann boundary condition on $\frac{\partial \mathbf{v}}{\partial n}$. To this end, we can either directly impose $\frac{\partial u_{1D}}{\partial x}$ or compute a linear combination (flux upwinding) $f(\frac{\partial \mathbf{v}}{\partial n}, \frac{\partial u_{1D}}{\partial x})$ where time-extrapolated values can also be considered:

$$\frac{\partial \mathbf{v}^{n+1}}{\partial n} = f\left(\frac{\partial \mathbf{v}}{\partial n}, \frac{\partial u_{1D}}{\partial x}\right) = \sum_{k=0}^{Je-1} \xi_k \left(\lambda \frac{\partial \mathbf{v}^{n-k}}{\partial n} - (1 - \lambda) \frac{du_{1D}^{n+1}}{dx} \right) \quad (5.51)$$

where $0 \leq \lambda \leq 1$. Typically, we choose $\lambda = 0.5$ (central averaging).

5.5.2.3 Arterial wall stiffness of the 1D model

Since we consider the coupling between rigid 3D and compliant 1D domains, an inconsistency arises at the coupling interface, where the continuity of cross sectional area is violated. Comparing the solutions computed from the two different models (3D vs 1D) given an identical simulation setup, we notice that while we can get a good match for the flow-rate, the predicted solutions for the propagating pressure wave are completely different. This is of course a well expected behavior and it is attributed to the different arterial wall response (rigid vs compliant). As first put forth in [Grinberg et al. \(2011a\)](#), a straightforward way for the compliant 1D model to reproduce the results of the rigid 3D solver is to make the arterial wall model stiffer by scaling appropriately the spring constant β that appears in the 1D pressure-area relation (see Eq. 5.3):

$$\beta = \beta_{scal} \frac{\sqrt{\pi} E h}{(1 - \nu^2) A_0}, \quad \beta_{scal} > 1 \quad (5.52)$$

Practically, we can find an appropriate scaling factor β_{scal}^* such that the 1D solver accurately reproduces the results predicted by the full 3D model. However, if the chosen value for β_{scal} is not close to the “best fit” value β_{scal}^* , the 1D solver under predicts the pressure wave for $\beta_{scal} \ll \beta_{scal}^*$ and over-predicts it for $\beta_{scal} \gg \beta_{scal}^*$.

The identification of the appropriate scaling factor is case depended. However, we can explicitly derive a formula for β_{scal}^* based on the physiological properties of the network. To this end, consider a single vessel in an arterial network and let R_t be the total peripheral resistance downstream, and R_1 the vessel's characteristic impedance:

$$R_1 \doteq \frac{c_0 \rho}{A_0}, \quad c_0 = \sqrt{\frac{1}{2} \frac{\beta}{\rho} A_0^{\frac{1}{4}}} \quad (5.53)$$

Then, we must have that $R_t - R_1 \geq 0$, and therefore, we can scale β and stiffen the 1D model respecting the upper bound:

$$R_1 \leq R_t \Rightarrow \beta_{scal} \leq \frac{2R_t^2 A_0^{\frac{3}{2}}}{\beta \rho} \quad (5.54)$$

In simulations, the total resistance R_t is typically prescribed for each outlet that is attached to a 0-D model and its values can be obtained either by tuning the model to match patient specific data, or, more generally, by interpolating available data sets in the literature (see for e.g., [Reymond et al. \(2011\)](#)).

Here, we also propose an alternative way for stiffening the wall response of the 1D model. To this end, we introduce a scaling factor $\tilde{\tau}$ to the 1D pressure-area relation as:

$$p = p_{ext} + \beta(\sqrt{A} - \sqrt{A_0}) \cdot \tilde{\tau} \quad (5.55)$$

where the values of $\tilde{\tau}$ are computed using the following low pass RC-filter analogue:

$$\tilde{\tau} + RC \frac{\partial \tilde{\tau}}{\partial t} = 1 + R \cdot \left\{ \beta(\sqrt{A} - \sqrt{A_0}) \right\} \quad (5.56)$$

Obviously, for $R = 0$ we recover the fully elastic 1D pressure-area relation, while for $R > 0$, we get $\tilde{\tau} > 1$, resulting to a reduced distensibility of the arterial wall cross section. The main advantage of this approach is that it achieves an accurate fit to the solution predicted by the full 3D model when we a-priori choose $R = F(t)\beta\sqrt{A}$, where $F(t) = (1 - e^{-20(t-t_0)})^2$ is a filter function used to filter out any erroneous oscillations that occur during the first time-steps. This choice is based on the fact that the factor $\beta\sqrt{A}$ eliminates the vessel's distensibility, which for a vessel governed by the 1D elastic pressure-area relation is defined by: $D \doteq \frac{1}{A} \frac{dA}{dp} = \frac{2}{\beta\sqrt{A}}$. Finally, the capacitance in the RC-filter equation for $\tilde{\tau}$ is typically set as $C = \frac{0.18}{\beta\sqrt{A_0}}$.

5.6 Results

In this section we append some representative results for our 3D-1D weak coupling algorithm. We start the section with a basic test case of Womersley flow in a straight pipe in order to compare the results obtained by the full 3D/1D models and highlight the different results obtained between the rigid 3D and the compliant 1D case. Next, using the same straight pipe setup, we truncate the 3D domain and couple its outlet to a 1D segment. The results presented aim to answer the question of how stable and accurate are the predictions given by our coupled 3D-1D solver when compared against the full 3D case. The same comparison is also carried out for a slightly more challenging configuration: a Y-shaped bifurcation geometry with two outlets.

5.6.1 3D/1D comparison in a straight pipe

Here we consider a straight pipe of length $L = 2\text{cm}$, and radius $r = 3\text{mm}$. For the 3D model impose the exact Womersley solution as an inflow boundary condition on the velocity, while a lumped parameter RC network ($R = 100$, $C = 0.002$) provides the Dirichlet outflow boundary condition for the pressure. Similarly, the 1D model is driven by the same Womersley solution, while we use a three-element windkessel model (RCR) that exactly matches the outflow boundary condition of the 3D setup. The Womersley number of the flow is $Ws = 0.5D\sqrt{\omega/\nu} = 5.585$, the Reynolds number is computed as $Re = \frac{U_m D}{\nu} = 135$ and the period of each cycle is $T = 0.423\text{s}$. The 3D computational domain is discretized by $N_{el} = 14245$ spectral elements, while the 1D domain consists of only 1 DG element. In both cases the solution is represented within each element using polynomial expansions of 4^{th} order while time-integration is 2^{nd} order accurate. Finally, the 3D simulation was performed using 1024 processors on a Cray XK6 machine (ORNL Jaguar).

Regarding the arterial wall stiffness of the 1D model, three cases are considered: 1.) fully elastic model, where β is set according to the data published by [Olufsen \(1999\)](#), 2.) stiff elastic model (denoted as β_{stiff}), where the initial β has been scaled to $1.4 \cdot 10^4 \beta$ (setting $\beta_{scal} = 1.4 \cdot 10^4$ results to a good fit of the 3D predictions), and 3.) stiff elastic model, where the scaling factor $\tilde{\tau}$ is computed using the RC-filter equation introduced in Sec. [5.5.2.3](#) (denoted as $\beta - RC$).

	3D	1D elastic	1D β_{stiff}	1D $\beta - RC$
Δt	0.002	0.06	0.005	0.005
Computation time/cycle	1 hr	$\ll 1\text{ s}$	20 s	20 s

Table 5.1: *3D/1D comparison in a straight pipe*: Non-dimensional time-step Δt and computation time per cycle for each simulation considered.

As we can see in the following figures, the 1D model captures the flow-rate wave for any configuration of the arterial wall model used. However, although the average pressure over the cycle is accurate, the actual pressure waveform predicted by the fully elastic 1D model is inconsistent with the solution in the 3D domain.

For a better fit we need to stiffen the 1D wall model. To this end, choosing $\beta_{scal} = 1.4 \cdot 10^4$ results to a good approximation, respecting the upper bound presented in Sec. 5.5.2.3, Eq. 5.54. Moreover, when $\beta_{scal} \ll \beta_{scal}^*$ or $\beta_{scal} \gg \beta_{scal}^*$ we get under/over-shoots in the predicted waveform as demonstrated in Fig. 5.9 for $\beta_{scal} = 10^3$ and $\beta_{scal} = 5 \cdot 10^4$.

Finally, choosing to stiffen the 1D model by using the $\beta - RC$ approach seems more appealing since accurate results are obtained without any parameter calibration needed.

Table 5.2, summarizes the relative $\mathbb{L}_2$ errors computed by comparing each of the considered 1D simulations to the full 3D case.

Relative $\mathbb{L}_2$ error	1D elastic	1D β_{stiff}	1D $\beta - RC$
$\frac{\ Q_{3D} - Q_{1D}\ _2}{\ Q_{3D}\ _2}$	5.001895e-02	1.105535e-02	1.080173e-02
$\frac{\ p_{3D} - p_{1D}\ _2}{\ p_{3D}\ _2}$	6.203276e-01	1.836266e-02	4.339720e-02

Table 5.2: *3D/1D comparison in a straight pipe*: Relative $\mathbb{L}_2$ error at the outlet.

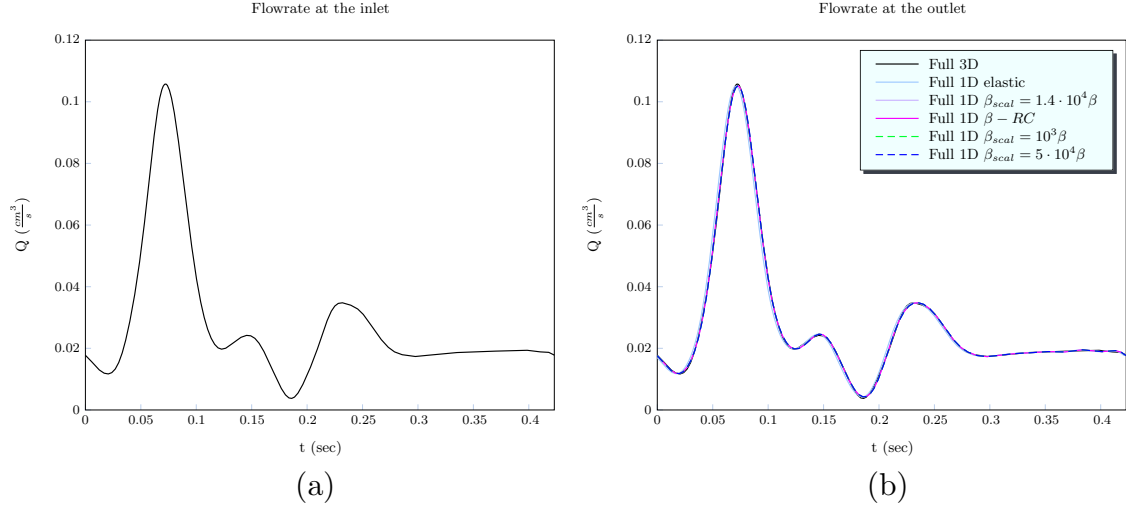


Figure 5.8: *3D/1D comparison in a straight pipe*: Comparison of computed flow-rates at the inlet and outlet.

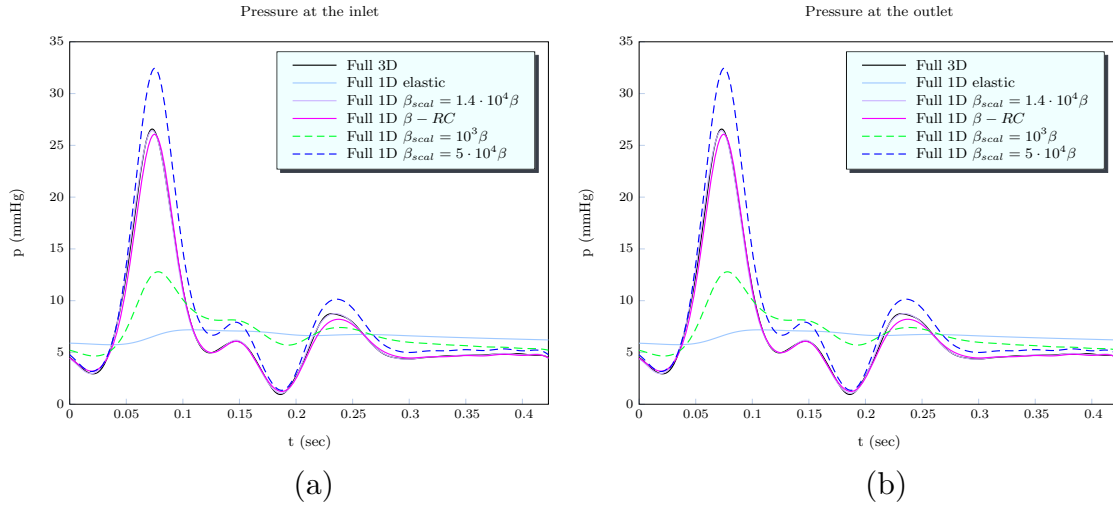


Figure 5.9: *3D/1D comparison in a straight pipe*: Comparison of computed pressures at the inlet and outlet.

As we can see, the elastic 1D model returns inconsistent predictions for the pressure. The best match is achieved when we choose to stiffen the arterial wall response of the 1D model using a scaling factor close to the one predicted by Eq. 5.54 in Sec. 5.5.2.3. This results to a relative error less than %2, while using the $\beta - RC$ stiffening approach results to a slightly higher error for the pressure (%4.3).

5.6.2 3D-1D coupling for a straight pipe

This is a basic test case for analyzing the accuracy of our coupling algorithm. We consider exactly the same setup presented in Section. 4.1, but now the 3D domain is truncated at $z = 1.4cm$ and the remaining portion of the pipe is replaced by an overlapping or non-overlapping 1D segment of length $l = 12mm$ or $l = 6mm$, respectively. In the overlapping case we choose the overlap length to be equal to twice the radius $l_{overlap} = 2r = 6mm$. The set of 3D-1D coupling conditions considered corresponds to cases 1,2 presented in Section 3 and will be denoted as p -coupling and RC -coupling, respectively. Regarding the arterial wall stiffness of the 1D model, we consider the same three cases presented above: 1.) fully elastic model, where β is set according to the data published by Olufsen (1999), 2.) stiff elastic model (denoted as β_{stiff}), where the initial β has been scaled to $1.533 \cdot 10^4 \beta$ (using Eq. 5.54, Eq. 5.54), and 3.) stiff elastic model, where the scaling factor $\tilde{\tau}$ is computed using the RC-filter equation introduced in Sec. 5.5.2.3 (denoted as $\beta - RC$). The truncated 3D computational domain consists of $N_{el} = 9908$ spectral elements and the computation time for 1 cycle on 1024 processors (Cray XT5) was $37mins$; a %40 speedup compared to the full 3D case. The time step of the 3D solver was $\Delta t_{3D} = 0.002$ non-dimensional time units, while the 1D solver was advanced in time using $\Delta t_{1D} = \Delta t_{3D}/5$ for stability reasons.

Results are presented for the following cases: 1.) Full 3D domain, 2.) 3D-1D coupling with a fully elastic pressure-area relation (β is set according to the data published by Olufsen (1999)), 3.) 3D-1D coupling with a stiff 1D model (denoted as β_{scal}^* , with the fixed scaling factor $\beta_{scal}^* = 1.533 \cdot 10^4$ for a good fit of the 3D predictions), and, 4.) 3D-1D coupling with a stiff 1D model (denoted as $\beta - RC$, where the scaling factor $\tilde{\tau}$ is computed using the RC-filter equation introduced in

Eq. 5.54). Using $\beta_{scal} \ll \beta_{scal}^*$ or $\beta_{scal} \gg \beta_{scal}^*$ we observe again under/over predictions in the pressure wave exactly similar to the ones presented in Sec. 5.6.1 so we will omit presenting them again here.

The coupling interface conditions used correspond to cases 1,2 (see Sec. 5.5), i.e the pressure at the 3D outflow is either imposed directly from the 1D domain or through an RC filter (denoted as p -coupling and RC -coupling, respectively).

Table 5.3 summarizes the computed errors for each case considered.

		RC -coupling			p -coupling
	Relative $\mathbb{L}_2$ error	elastic	β_{scal}^*	$\beta - RC$	β_{scal}^*
No overlap	$\frac{\ Q_{3D} - Q_{1D}\ _2}{\ Q_{3D}\ _2}$	5.001895e-02	1.518175e-02	1.525898e-02	1.518176e-02
	$\frac{\ p_{3D} - p_{1D}\ _2}{\ p_{3D}\ _2}$	6.203276e-01	1.802136e-02	4.413450e-02	1.802138e-02
Overlap	$\frac{\ Q_{3D} - Q_{1D}\ _2}{\ Q_{3D}\ _2}$	5.001213e-02	1.413325e-02	1.512868e-02	1.413327e-02
	$\frac{\ p_{3D} - p_{1D}\ _2}{\ p_{3D}\ _2}$	6.073351e-01	1.496531e-02	4.302153e-02	1.496537e-02

Table 5.3: *3D-1D coupling*: 3D-1D coupling for a straight pipe: Relative $\mathbb{L}_2$ error at the outlet.

As expected, the rigid 3D - elastic 1D coupling, although stable, yields inaccurate predictions for the pressure wave propagation. Only by 'stiffening' the wall response of the 1D model we can achieve a good fit to the full 3D case. In particular, both methods (β_{scal}^* , $\beta - RC$) where stable, with β_{scal}^* resulting to better accuracy. The stiff 1D cases require a smaller time-step than the elastic 1D model but this does not have any significant impact on the overall computation cost of this simulation.

Moreover, we observe agreement in the results for either set of interface coupling conditions. However, the RC -coupling seems more preferable since it was less com-

putationally expensive than p -coupling (by a factor of about 10% per cycle). This difference may be attributed in the 'filtered' Dirichlet pressure boundary condition imposed at the 3D outlet, which potentially helps reduce the number of iterations of the CG pressure solver.

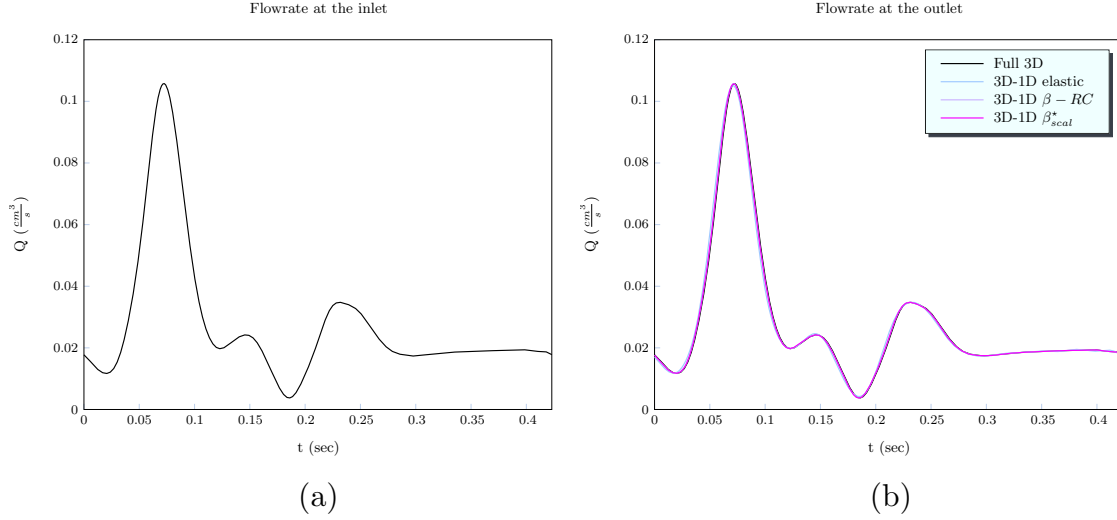


Figure 5.10: *3D-1D coupling*: 3D-1D coupling for a straight pipe: Comparison of computed flow-rates at the inlet and outlet (RC -coupling, no overlap).

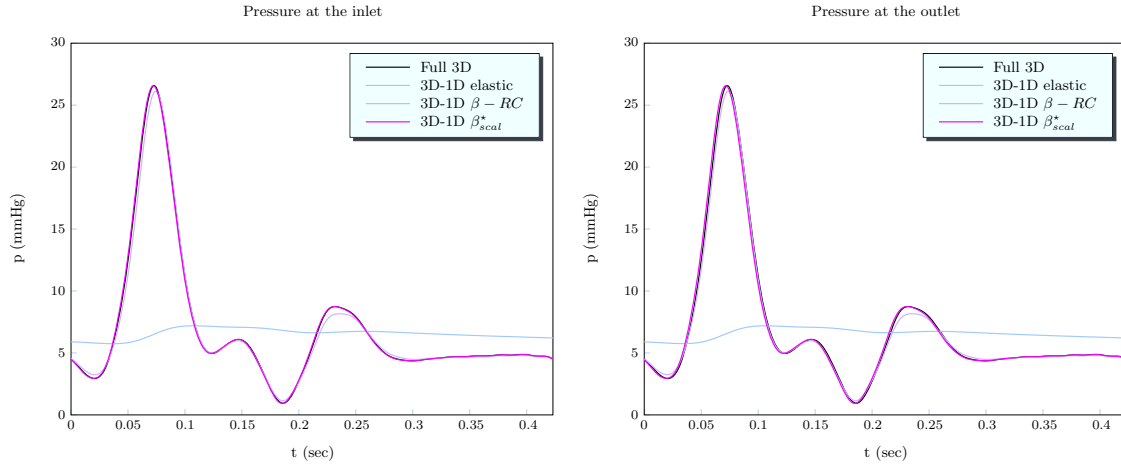


Figure 5.11: *3D-1D coupling*: 3D-1D coupling for a straight pipe: Comparison of computed pressures at the inlet and outlet (RC -coupling, no overlap).

The results are in agreement to what we saw in the full 3D vs full 1D comparison. Again, the elastic 1D model returns inconsistent predictions for the pressure. The

best match is achieved when we choose to stiffen the arterial wall response of the 1D model using the scaling factor predicted by Eq. 5.54 in Sec. 5.5.2.3. Again, this results to a relative error less than %2, while using the $\beta - RC$ stiffening approach results to a slightly higher error for the pressure (%4.3).

Both the non-overlapping and overlapping configurations return consistent results, with the later demonstrating a marginally better accuracy. Although, all coupling configurations where stable for this case, this may not be the case in general. As we will see in the next section, stability issues arise when a slightly more challenging test case is addressed.

5.6.3 3D-1D coupling for a Y-shaped bifurcation

We consider a Y-shaped bifurcation domain with one inlet and two outlets. The dimensions of each segment of the full 3D domain are: $\{l_1 = 10mm, r_1 = 3mm\}$, $\{l_2 = 25mm, r_2 = 3mm\}$, $\{l_3 = 20mm, r_3 = 1.5mm\}$. The simulation is driven by prescribing a Womersley inflow at the inlet with $Ws = 0.5D\sqrt{\omega/\nu} = 5.585$, $Re = \frac{U_m D}{\nu} = 135$, and $T = 0.423s$. The constant mode of the Womersley solution is scaled so that the magnitude of the imposed flow-rate lies in the physiological range for vessels of this size. For the full 3D reference case, the RC outflow boundary condition is used, and the RC parameters are chosen such that the flow-rate ratio at the two outlets is $Q_1/Q_2 = 13$. The computational domain for the full 3D case consists of $Nel = 45441$ spectral elements with the polynomial expansion within each element being of 4^{th} order. Each cycle took approximately *73mins* on 2048 cores of IBM BlueGene/P (Intrepid ANL) using a non-dimensional time-step $\Delta t_{3D} = 0.002$ and 2^{nd} order time integration.

For the 3D-1D coupling, the original domain is truncated such that: $\{l_1 = 10mm, r_1 = 3mm\}$, $\{l_2 = 15mm, r_2 = 3mm\}$, $\{l_3 = 15mm, r_3 = 1.5mm\}$, with the remaining topology being represented by 1D domains. This truncated 3D domain is then discretized in space using $N_{el} = 29106$ spectral elements with 4^{th} polynomial expansions within each element. Similarly, the 1D domains consist of 1 (non-overlapping case) or 2 (overlapping case) DG elements of the same order. For the overlapping case, the 3D-1D overlap length was set equal to twice the radius: $l_{overlap}^1 = 2r_1 = 6mm$ and $l_{overlap}^2 = 2r_2 = 3mm$. To ensure the stability of the 1D solver we use 5 inner time-steps with $\Delta t_{1D} = \Delta t_{3D}/5$.

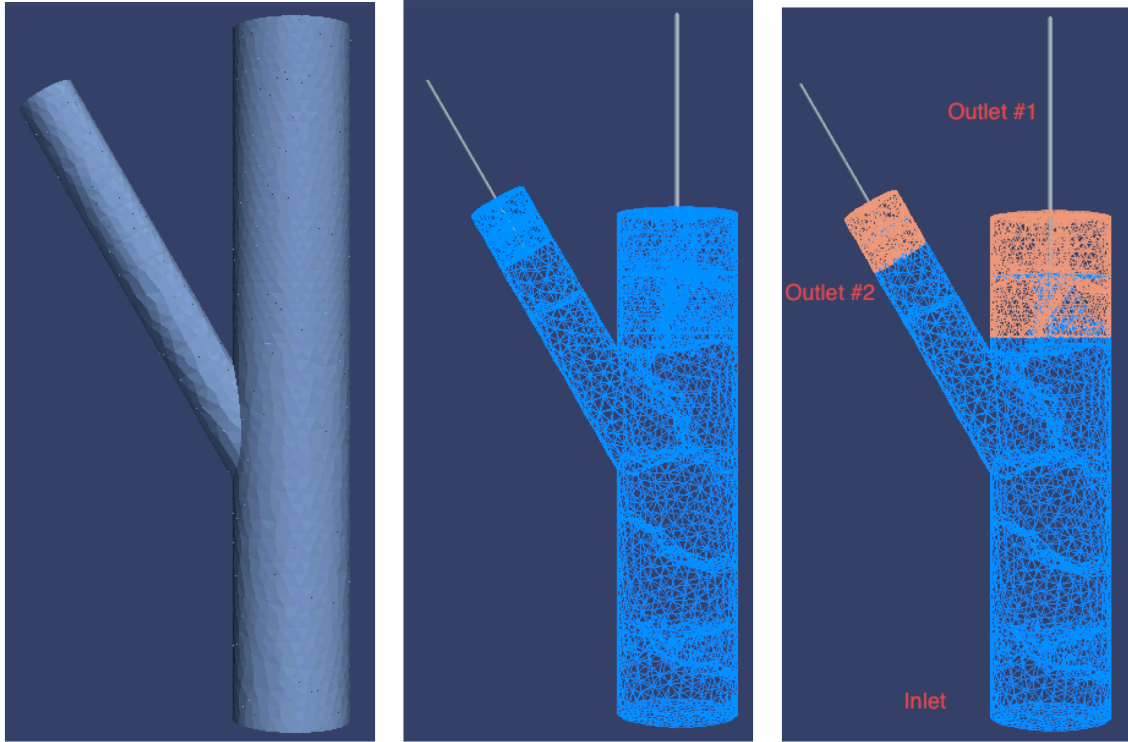


Figure 5.12: *3D-1D coupling*: 3D-1D coupling for a Y-shaped bifurcation: Considered model for the Y-shaped bifurcation: full 3D domain (left), 3D-1D non-overlapping domain (center), 3D-1D overlapping domain (right). The orange region depicts the overlap of the 3D and 1D domains.

The configurations considered are: 1.) Full 3D domain, 2.) 3D-1D coupling with a fully elastic pressure-area relation (β is set according to the data published by

Olufsen (1999)), 3.) 3D-1D coupling with a stiff 1D model (denoted as β_{scal}^* , using the scaling factor presented in Eq. 5.54 in Sec. 5.5.2.3), and, 4.) 3D-1D coupling with a stiff 1D model (denoted as $\beta - RC$, where the scaling factor $\tilde{\tau}$ is computed using the RC-filter equation introduced in Sec. 5.5.2.3).

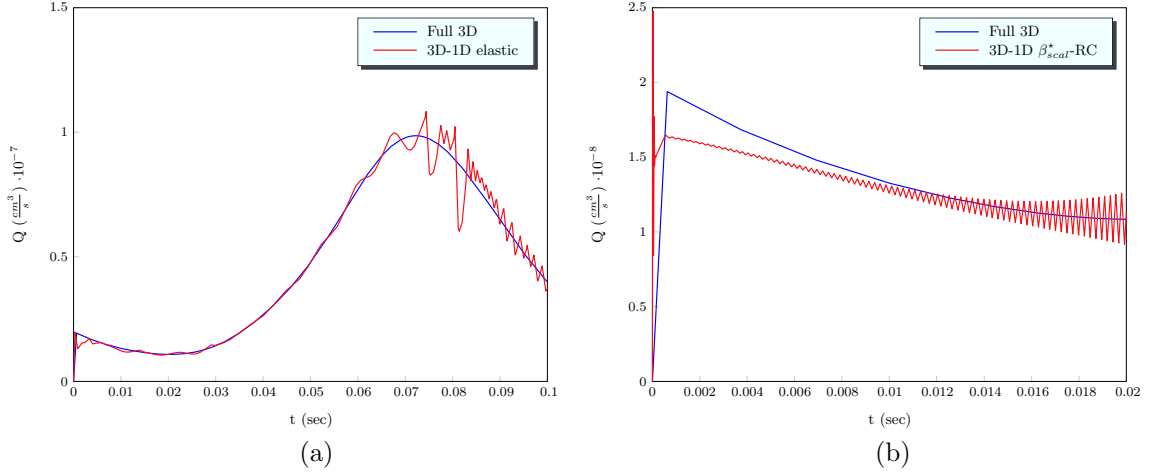


Figure 5.13: *3D-1D coupling*: 3D-1D coupling for a Y-shaped bifurcation: Flow-rate oscillations at the 3D-1D non-overlapping interface #1 (*RC*-coupling).

Unlike the straight pipe case where all configurations and models were stable, here we encounter issues that render all the non-overlapping coupling cases to be unstable. Wave reflections at the 3D-1D interfaces may contaminate the solution and amplify high modes, inducing oscillations in the flow-rate and pressure waves. These instabilities are manifested in an explosive manner in the case of stiff 1D models, leading to an immediate blow-up during the first hundred time-steps (see Fig. 5.13(a)). In the case of elastic 1D domains, the instability appears milder but still the solution diverges at some point during the first quarter of the cycle (a few thousand time-steps, see Fig. 5.13(b)). Moreover, the instabilities are even stronger when *p*-coupling is used instead of *RC*-coupling, even if we try to filter the initial oscillations due to non divergence free initial conditions.

		<i>RC</i> -coupling			<i>p</i> -coupling
	Relative $\mathbb{L}_2$ error	elastic	β_{scal}^*	$\beta - RC$	β_{scal}^*
No overlap	$\frac{\ Q_{3D} - Q_{1D}\ _2}{\ Q_{3D}\ _2}$	Unstable	Unstable	Unstable	Unstable
	$\frac{\ p_{3D} - p_{1D}\ _2}{\ p_{3D}\ _2}$	Unstable	Unstable	Unstable	Unstable
Overlap	$\frac{\ Q_{3D} - Q_{1D}\ _2}{\ Q_{3D}\ _2}$	1.834868e-01	1.444825e-02	Unstable	Unstable
	$\frac{\ p_{3D} - p_{1D}\ _2}{\ p_{3D}\ _2}$	7.080739e-01	3.316665e-02	Unstable	Unstable

Table 5.4: *3D-1D coupling*: 3D-1D coupling for a Y-shaped bifurcation: Relative $\mathbb{L}_2$ error at outlet #1.

		<i>RC</i> -coupling			<i>p</i> -coupling
	Relative $\mathbb{L}_2$ error	elastic	β_{scal}^*	$\beta - RC$	β_{scal}^*
No overlap	$\frac{\ Q_{3D} - Q_{1D}\ _2}{\ Q_{3D}\ _2}$	Unstable	Unstable	Unstable	Unstable
	$\frac{\ p_{3D} - p_{1D}\ _2}{\ p_{3D}\ _2}$	Unstable	Unstable	Unstable	Unstable
Overlap	$\frac{\ Q_{3D} - Q_{1D}\ _2}{\ Q_{3D}\ _2}$	3.745957e-01	2.403443e-02	Unstable	Unstable
	$\frac{\ p_{3D} - p_{1D}\ _2}{\ p_{3D}\ _2}$	7.025703e-01	4.286251e-02	Unstable	Unstable

Table 5.5: *3D-1D coupling*: 3D-1D coupling for a Y-shaped bifurcation: Relative $\mathbb{L}_2$ error at outlet #2.

The only way we could obtain both stable and accurate results was by using the 3D-1D *RC*-coupling with an overlapping region between the coupled domains. As we can see in Tables 5.4, 5.5 and Figs. 5.14, 5.15, the relative error in the flow-rate and the pressure between the full 3D and 3D-1D coupling cases was less than 5% when the 1D vessels were stiffened according to Eq. 5.54 in Sec. 5.5.2.3 (case β_{scal}^*). Interestingly, the overlapping coupling between the rigid 3D domain and the purely elastic 1D domain was initially stable but then starts diverging after the second cycle (Fig. 5.14(c)).

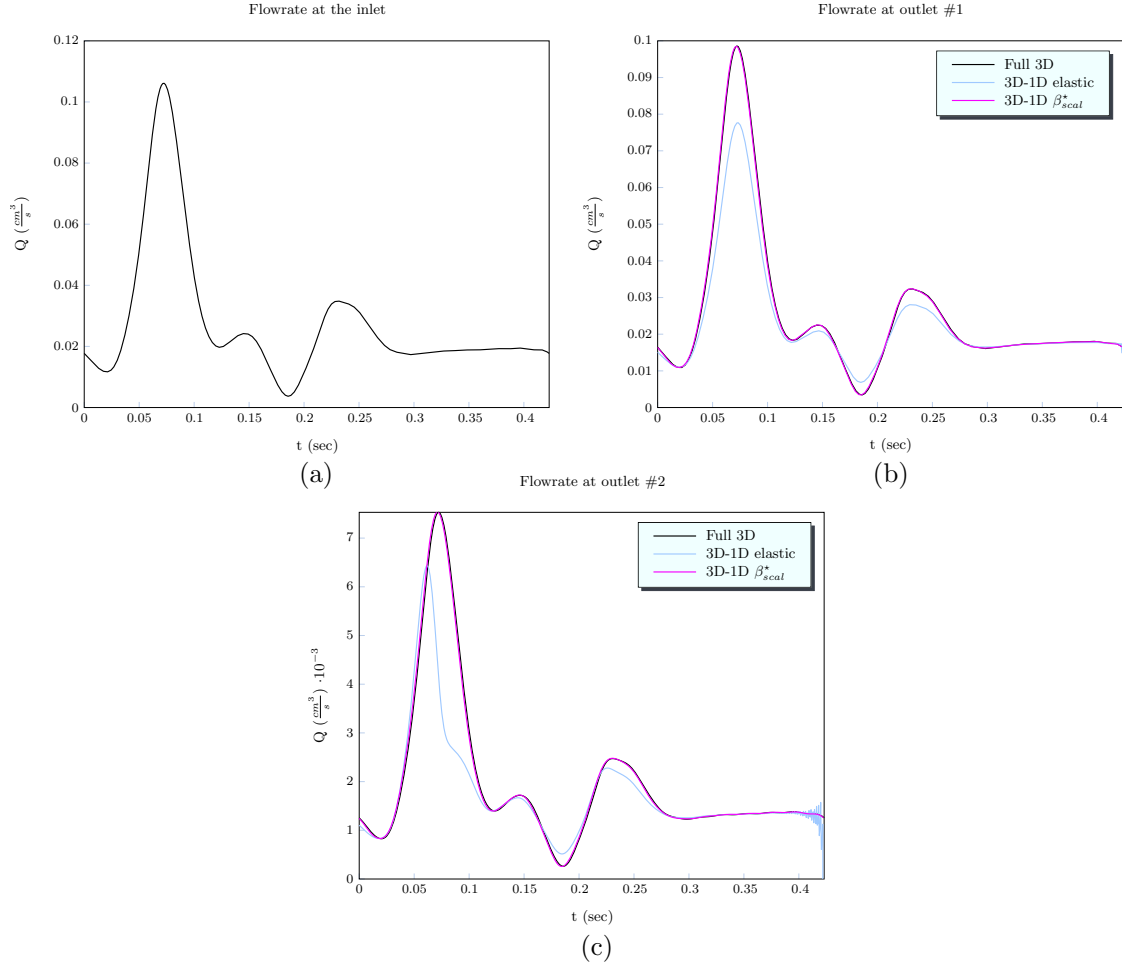


Figure 5.14: *3D-1D coupling*: Comparison of computed flow-rates at the inlet and outlets (*RC*-coupling, 3D-1D overlap).

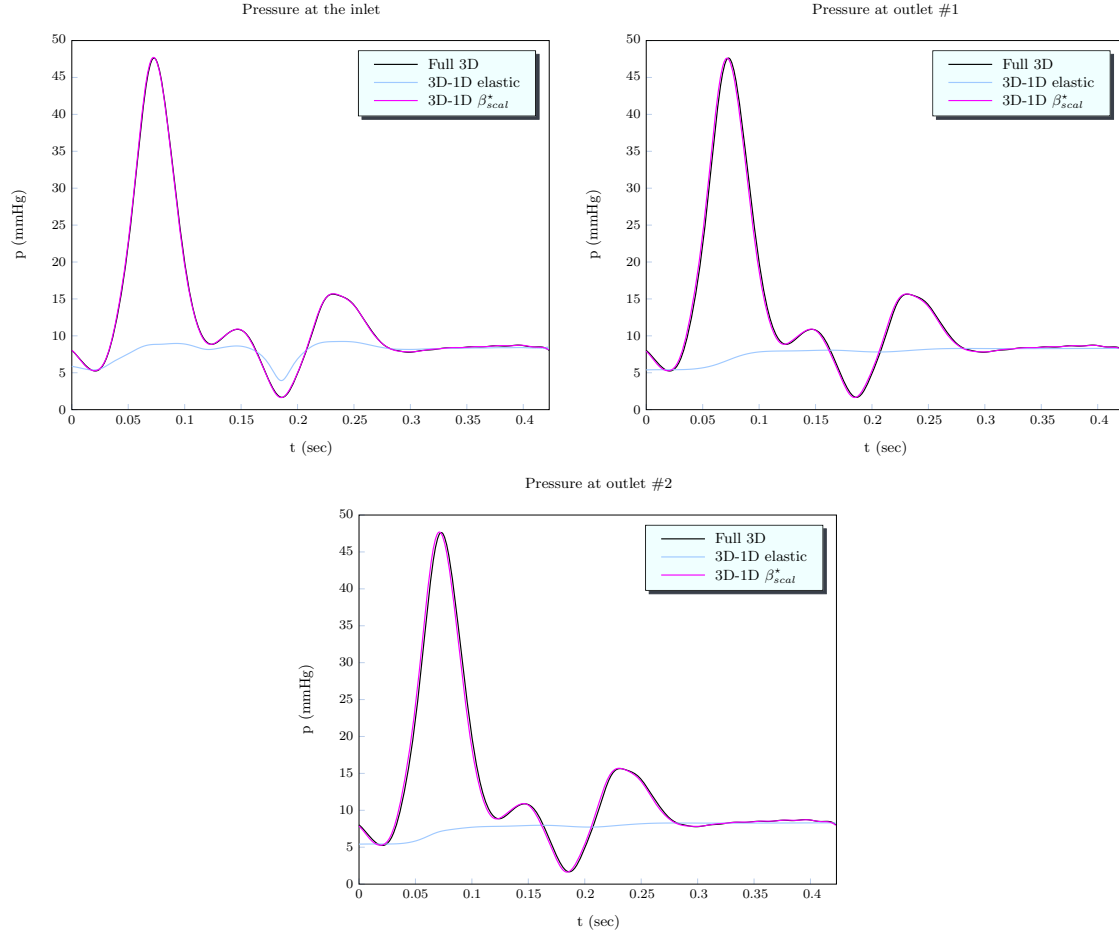


Figure 5.15: *3D-1D coupling*: Comparison of computed pressures at the inlet and outlets (*RC*-coupling, 3D-1D overlap).

CHAPTER SIX

Fractional-order Viscoelasticity in Arterial Biomechanics

6.1 Motivation

The vascular wall is a heterogeneous soft tissue with complex bio-mechanical properties that vary over different locations in the arterial tree and the temporal state of the response (deformation/relaxation). It is comprised of three different layers: the intima (mainly endothelial cells), the media (mainly smooth muscle cells), and the adventitia (mainly collagenous fibers) (see [Fung \(1990\)](#)). This structure highlights two key functions, namely, reversible energy storage (elasticity) and energy dissipation (viscosity).

Constitutive laws for biological tissue can be derived using integer-order differential equations that model stress-strain relations using additive combinations of purely elastic and viscous elements (see [Fung \(1990\)](#)). The simplest integer-order models of linear viscoelasticity are the Voigt and Standard Linear Solid (SLS or Kelvin-Zener) models (see [Eringen \(1980\)](#)). The Voigt model can be constructed by the parallel combination of a spring and a dashpot and it is the simplest model that accounts for creep and hysteresis phenomena. Similarly, the SLS model is constructed using the parallel combination of a spring with a spring and a dashpot in series, and it accounts for creep, hysteresis and stress relaxation phenomena.

Due to their simplicity, these models have been used in several 1D blood flow studies to describe arterial wall viscoelasticity (see [Matthys et al. \(2007\)](#), [Reymond et al. \(2009\)](#), [Steele et al. \(2011\)](#), [DeVault et al. \(2008\)](#), [Čanić et al. \(2006\)](#), [Raghu et al. \(2011\)](#)). [Matthys et al. \(2007\)](#) employed a Voigt viscoelastic model within a 1D flow solver and validated results against *in vitro* measurements of flow in 37 silicone branches representing the largest central systemic arteries in the human. In [Reymond et al. \(2009\)](#), a more generalized integer-order viscoelastic model is considered,

and 1D simulations are compared against *in vivo* measurements in a patient-specific network of 94 large systemic arteries. Both studies show good agreement between simulations and measured data, indicating that 1D models may capture the most significant features of the pulsatile flow and pressure waveforms. However, they mainly focus on larger systemic arteries, where the arterial wall response is known to be predominantly elastic (see [Craiem and Armentano \(2007\)](#), [Craiem et al. \(2008\)](#)), whereas the effect of wall viscoelasticity on hemodynamics is minor and only manifested in distal locations (see [Reymond et al. \(2009\)](#)), where the arterial wall appears to be more muscular.

The theory of quasilinear viscoelasticity (QLV) theory (see [Fung \(1990\)](#)) provides a path to constructing integer-order models that accurately capture the time-dependent response of the arterial wall; however, estimating the elastic and viscoelastic parameters that define these models in different anatomic locations still remains a very challenging task. In particular, one needs to address the lack of clinical data, the local and patient-specific nature of parameter values, the dynamic variation that parameters such as the Young's modulus E may exhibit during the cardiac cycle, etc. Also, as reported by [Doehring et al. \(2005\)](#), the sensitivity on these parameters (especially the long viscoelastic relaxation time) is high. These are key limiting factors for considering integer-order QLV models in 1D blood flow simulations of large arterial networks with high spatial variability in mechanical properties and presence of regions where the viscoelastic response of the wall may be significant.

An alternative approach to modeling the viscoelastic behavior of arteries based on *fractional calculus* (see [Mainardi \(2010\)](#), [Podlubny \(1998\)](#), [Magin \(2006\)](#)) has been introduced by [Craiem and Armentano \(2007\)](#), [Craiem et al. \(2008\)](#), [Doehring et al. \(2005\)](#), where the cell and tissue biomechanics are modeled by *fractional-order* differential equations. Under this framework, a spring represents a zero-order element

(purely elastic response) and a dashpot corresponds to a first-order element (purely viscous response). A fractional order $0 < \alpha < 1$ results in a new flexible element type, the so-called *spring-pot* (see [Magin \(2006\)](#)), which essentially interpolates between the purely elastic and viscous behaviors and can be thought of as a large set of weighted integer-order spring/dashpot pairs arranged in parallel (see [Näsholm and Holm \(2013\)](#)). This gives rise to a very interesting physical interpretation: the fractional order introduces a fractal tree-like hierarchical structure, naturally allowing for separation of material and global time scales (see [Doehring et al. \(2005\)](#), Figure 1 therein). In contrast to the discrete relaxation spectrum of linear Voigt and SLS integer-order models, fractional-order models exhibit a continuous relaxation distribution, spanning all frequencies above zero, up to infinity. Therefore, fractional-order models can be naturally put under the context of quasi-linear viscoelasticity and be considered as good candidates for modeling biological tissue, which in reality shows continuous relaxation (see [Fung \(1990\)](#)).

Recent experimental studies have confirmed the ability of fractional-order models to capture the dynamic response of the vascular wall. Specifically, [Craiem and Armentano \(2007\)](#) proposed a fractional-order Voigt model (FO-Voigt) and accurately described the viscoelastic mechanical response of the aorta of a sheep *in-vivo* with a minimal set of estimated parameters. Also, [Craiem et al. \(2008\)](#) performed uniaxial loading experiments on strips cut from healthy human aortas, and a FO-Voigt model was successfully fitted to measured stress-relaxation data. Similarly, [Craiem et al. \(2008\)](#) calibrated FO-Voigt and fractional-order SLS (FO-SLS) models to accurately reproduce *in-vivo* data from ascending aorta segments of four human donors. [Doehring et al. \(2005\)](#) integrated a fractional stress relaxation function corresponding to a FO-SLS model in the QLV framework to successfully model aortic valve cusp biomechanics, while demonstrating significantly lower sensitivity on the input

parameters as compared to a standard QLV model with a continuous relaxation spectrum. These recent findings indicate that fractional order models may offer a new powerful alternative for describing arterial wall mechanics, reducing the parameter estimation count and overcoming the sensitivity limitations of integer-calculus-based QLV models.

In this chapter, we present a nonlinear, time-domain 1D blood flow formulation integrated with integer- and fractional-order viscoelastic wall models. We perform blood flow simulations in a large patient-specific network and compare integer- and fractional-order SLS models, calibrated with experimental data reported in the literature. Our main scientific goal is to quantify the effect of the fractional-order and the viscoelastic relaxation response on the computed flow and pressure wave propagation. Finally, we present a global sensitivity analysis study for which we have considered a stochastic fractional-order SLS model by introducing uncertainty in the parameters that define its viscoelastic behavior.

6.2 Viscoelastic Models of the Arterial Wall

The theory of quasi-linear viscoelasticity (see [Fung \(1990\)](#)) provides a unifying framework under which general soft tissue constitutive laws can be formulated. Such models are often characterized by a constant hysteresis over a wider frequency range. Based on this approach we have a stress-strain relation of the form

$$\sigma(x, t) = \int_0^\infty G(t - \gamma) \frac{\partial \epsilon^e(x, \gamma)}{\partial \gamma} d\gamma, \quad (6.1)$$

where $G(t)$ is the stress relaxation function and $\epsilon^e(x, t)$ is the static elastic response of the tissue. The presence of the convolution integral in Eq. 6.1 makes the stress depend upon the strain time history, but it is the choice of the stress relaxation function that ultimately determines the tissue model. Typically, one introduces a parametric representation of $G(t)$ that fits given experimental measurements (see [Steele et al. \(2011\)](#), [Valdez-Jasso et al. \(2011\)](#)). In the following sections we present the fractional-order Standard Linear Solid (SLS) model and its implementation in a one-dimensional blood flow solver.

6.2.1 The Fractional SLS model (Fractional Kelvin-Zener model)

Before we present the fractional version of the SLS model, we start with the classical definition of the Caputo fractional derivative of order α (see [Mainardi \(2010\)](#))

$${}_0^C D_t^\alpha f(t) := \begin{cases} \frac{1}{\Gamma(n - \alpha)} \int_0^t \frac{f^{(n)}(\tau)}{(t - \tau)^{\alpha+1-n}} d\tau, & n - 1 < \alpha < n \\ \frac{d^n f(t)}{dt^n}, & \alpha = n \end{cases}, \quad (6.2)$$

where $\alpha > 0$ is a real number, n is an integer, and $\Gamma(\cdot)$ is the Euler gamma function. We note that for $\alpha \neq n$, the Caputo derivative is a non-local operator that depends on the history of f in the interval $[0, t]$.

The fractional order generalization of the SLS model is constructed using the parallel combination of a spring with a spring in series with a *spring-pot*. The stress is related to strain as

$$\sigma(t) + \tau_{\sigma 0}^{\alpha C} D_t^\alpha \sigma(t) = E [\epsilon(t) + \tau_{\epsilon 0}^{\alpha C} D_t^\alpha \epsilon(t)] \quad (6.3)$$

We observe that simpler fractional- and integer-order viscoelastic models as well as the purely elastic tube law appear as limiting cases in the above equation. Namely, for $\alpha = 1$, we recover the classic integer-order SLS model, while taking $\tau_\sigma = 0$ results to the fractional-order Voigt model.

Thermodynamic consistency of the FO-SLS model requires a monotonically decreasing stress relaxation function, which further implies that the fractional time derivatives of both the stress and the strain must be of the same order α (see [Atanacković et al. \(2011\)](#), [Näsholm and Holm \(2013\)](#)).

By employing the Laplace transform, we can express the FO-SLS stress relaxation function as

$$G(t) = E \left(\frac{\tau_\epsilon}{\tau_\sigma} \right)^\alpha + E \left(\left[\frac{\tau_\epsilon}{\tau_\sigma} \right]^\alpha - 1 \right) E_{\alpha,1} \left(- \left[\frac{t}{\tau_\sigma} \right]^\alpha \right), \quad (6.4)$$

where $E_{\alpha,\beta}(t)$ is the two parameter Mittag-Leffler function, known as the fractional generalization of the exponential

$$E_{\alpha,\beta}(t) := \sum_{n=0}^{\infty} \frac{t^n}{\Gamma(\alpha - n\beta)} \quad (6.5)$$

Compared to the standard exponential function, the Mittag-Leffler function exhibits very different behavior [Mainardi \(2010\)](#): for small times it exhibits a much faster decay than the exponential, while at the limit $t \rightarrow \infty$ the decay is only algebraic. This heavy tail decay is what characterizes the Mittag-Leffler function as a *super-slow process* (see [Mainardi \(2010\)](#)).

Another important observation here is that for the FO-SLS model the idea of relaxation times must be revised. Under the fractional setting, they no longer represent single discrete relaxation frequencies but they should be rather be interpreted as

break-frequencies in a Cole-Cole model sense, around which the model characteristics change (see [Näsholm and Holm \(2013\)](#)).

Finally, by substituting the FO-SLS stress relaxation function of Eq. 6.4 in Eq. 6.1, integrating by parts and employing the definitions in Eq. 5.5 we arrive at the FO-SLS pressure-area relation

$$p(x, t) = p_{ext} + p^E(x, t) + p^V(x, t), \quad (6.6)$$

where p^E , p^V correspond to the elastic and viscoelastic pressure contributions, respectively.

$$\begin{aligned} p^E(x, t) &= \left(\frac{\tau_\epsilon}{\tau_\sigma} \right)^\alpha \beta(\sqrt{A} - \sqrt{A_0}) \\ p^V(x, t) &= \frac{1}{\tau_\sigma} \left(1 - \left[\frac{\tau_\epsilon}{\tau_\sigma} \right]^\alpha \right) \int_0^t E_{\alpha,0} \left(- \left[\frac{t-\gamma}{\tau_\sigma} \right]^\alpha \right) p^E(\gamma) d\gamma \end{aligned} \quad (6.7)$$

We observe that the FO-SLS model introduces stress-strain memory effects that in the long time limit decay algebraically due to the Mittag-Leffler relaxation kernel (see [Mainardi \(2010\)](#)). It is important to note here that thanks to the linearity of the time-fractional ODE governing the fractional-order SLS model (Eq. 6.3), the Laplace transform allows us to arrive to Eq. 6.6, which corresponds to the exact solution of Eq. 6.3.

An alternative way of formulating the required pressure-area relation arises from directly discretizing the Caputo time derivatives in the fractional constitute law that defines each model using an appropriate discretization technique. The easiest and most popular way of doing this is by employing the Grünwald-Letnikov formula (see

Podlubny (1998)):

$${}_0^C D_t^\alpha f(t) = \lim_{\Delta t \rightarrow 0} \Delta t^{-\alpha} \sum_{k=0}^{\infty} GL_k^\alpha f(t - k\Delta t), \quad GL_k^\alpha := \frac{k - \alpha - 1}{k} GL_{k-1}^\alpha, \quad (6.8)$$

with $GL_0^\alpha = 1$. By substituting the Grünwald-Letnikov formula in Eq. 6.3, we can formulate the pressure-area relation for the FO-SLS model as

$$p(x, t) = p_{ext} + \frac{1 + \tau_\epsilon^\alpha \Delta t^{-\alpha}}{1 + \tau_\sigma^\alpha \Delta t^{-\alpha}} p^E(x, t) + \frac{\Delta t^{-\alpha}}{1 + \tau_\sigma^\alpha \Delta t^{-\alpha}} \sum_{k=0}^{\infty} GL_k^\alpha \{ \tau_\epsilon^\alpha p^E(t - k\Delta t) - \tau_\sigma^\alpha p(t - k\Delta t) \}, \quad (6.9)$$

where the last term in the summation is the total elastic and viscoelastic pressure from previous time steps.

An attractive feature of this approach is that for a small time step Δt , which is typically the case for the high-order polynomial approximations employed here (due to the Courant-Friedrichs-Lewy (CFL) condition), the Grünwald-Letnikov coefficients exhibit fast decay properties. This enables us to reduce the computation of the convolution sum in Eq. 6.9, by using the “short memory” principle of Podlubny (1998), and approximate the viscoelastic memory effects using only a portion of the response history, disregarding any terms in the Grünwald-Letnikov expansion that are below a cutoff threshold. However, the elastic behavior of larger systemic arteries typically corresponds to low values of the fractional order (see Craiem and Armentano (2007), Craiem et al. (2008)), and the accurate evaluation of these convolutions using the “short memory” principle requires one to consider history effects from the last four cardiac cycles. Our numerical experiments indicate that this is the minimum amount of time-history required by the Grünwald-Letnikov formula to give numerically stable and convergent results for the problem considered. With our goal being the long-time integration of Eq. 5.1, using a time step as low as $\Delta t = 10^{-6}$, and with each cardiac cycle being about 1sec long, this results to storing at least

$4 \cdot 10^6$ values per history point. This value, in combination with the requirement to compute a convolution sum at each quadrature point and for each time step, renders the simulation completely *unrealistic* even on modern supercomputers. In Section 6.2.2, we will present a formulation that alleviates the computational complexity introduced by fractionality, and finally arrive at a computable workflow that employs the exact solution of Eq. 6.3 and accounts for the full time history in the evaluation of hereditary integrals.

It is important to note that viscoelasticity adds dissipation, leading to a system that is no longer strictly hyperbolic. However, the conservative elastic contribution is still dominant, thus allowing us to employ the Discontinuous Galerkin methodology presented in Sec. 5.2.2 without any numerical issues. We have observed that in cases in which the viscoelastic response is significant (integer-order models or fractional models with α close to 1.0) and the relaxation times ratio $\frac{\tau_\epsilon}{\tau_\sigma}$ becomes too large, the scheme becomes unstable. In such cases, due to the explicit nature of our method, stability is only recovered by decreasing the time-step sufficiently so that it resolves the viscous time scale. Alternatively, this issue could be alleviated by using a total variation diminishing (TVD) implicit time integration scheme (see Shu and Osher (1988)), however at the expense of increased computational cost and implementation complexity.

6.2.2 Evaluation of Hereditary Integrals

A major computational challenge in considering viscoelastic models is posed by the need to evaluate the convolution integral in the expression for the viscoelastic pressure component, $p^v(t)$. This evaluation quantifies the stress-strain memory dependencies and needs to be repeated at every time step, and for every quadrature point

of the domain. A naive implementation of this operation scales quadratically with the number of time steps, introducing a major computational bottleneck. Moreover, a typical 1D simulation of the hyperbolic system of Eq. 5.1, involving 50 arteries with a polynomial order of $P = 6$, may impose a very restrictive CFL condition, requiring a time step as low as $\Delta t = 10^{-6}$. This results to a very expensive computation both in terms of floating point operations and memory requirements. Here, we briefly present the methods and techniques we employed in this work to alleviate this issue and obtain a computable workflow.

Integer-order SLS model: For the SLS model we explore the properties of the exponential relaxation kernel and derive a simple updating rule that is independent of the time history imposed by the convolution operator:

$$I_{t+\Delta t} = \int_0^{t+\Delta t} e^{-\frac{(t+\Delta t-\gamma)}{\tau\sigma}} p^e(\gamma) d\gamma = e^{-\frac{\Delta t}{\tau\sigma}} I_t + \int_t^{t+\Delta t} e^{-\frac{(t+\Delta t-\gamma)}{\tau\sigma}} p^e(\gamma) d\gamma, \quad (6.10)$$

where the last term corresponds to the time interval $[t, t + \Delta t]$ and can be computed using standard quadrature rules. This effectively allows us to consider the full history of the arterial wall response, without introducing any computational or memory bottleneck, as only the solution from the last time step needs to be stored.

Fractional order models: Due to the presence of power law (FOV-Kelvin-Voigt model) or Mittag-Leffler (FOV-SLS) relaxation kernels, fractional order models do not admit the separability concept introduced above to obtain a simple updating rule for the hereditary integral. However, one can still avoid the “brute force” computation by appropriately choosing a smart algorithm. Here, we have employed the method developed by [Lubich and Schädle \(2002\)](#). Following their approach, we employ the definition of the inverse Laplace transform, to generally write a hereditary

integral in the time interval (a, b) as:

$$\begin{aligned} \int_a^b f(t - \gamma)g(\gamma)d\gamma &= \int_a^b \frac{1}{2\pi i} \left(\int_{\Gamma} F(s)e^{(t-\gamma)s} ds \right) g(\gamma)d\gamma \\ &= \frac{1}{2\pi i} \int_{\Gamma} F(s)e^{(t-b)s} \int_a^b e^{(t-\gamma)s} g(\gamma)d\gamma ds, \end{aligned} \quad (6.11)$$

where the inner integral, denoted as $y(b, a, s)$, is recognized as the solution at time b of the homogeneous scalar initial value problem:

$$y' = sy + g, \quad y(a) = 0 \quad (6.12)$$

Then, we can choose an integer base parameter B and discretize the global time interval in a local sequence of fast growing intervals, $I_l = [B^{l-1}\Delta t, 2(B^l - 1)\Delta t]$. Moreover, we can apply an N point quadrature rule for the numerical evaluation of contour integral in Eq. 6.11, and discretize the convolution operation as:

$$\int_a^b f(t - \gamma)g(\gamma)d\gamma \approx \sum_{l=1}^L \sum_{j=-N}^N w_j^{(l)} F(s_j^{(l)}) e^{(t-\gamma_{l-1})s_j^{(l)}} y(\gamma_{l-1}, \gamma_l, s_j^{(l)}), \quad (6.13)$$

where the nodes $s_j^{(l)}$ and the weights $w_j^{(l)}$ are determined by the choice of the quadrature rule employed to evaluate the contour integral in Eq. 6.11 for every approximation interval $[t - \gamma_{l-1}, t - \gamma_l] \subset I_l$. Here, we have chosen a trapezoidal rule with equidistant steps to a parametrized *Talbot contour* (see [Lubich and Schädle \(2002\)](#)), yet we note that other choices of contours, such as hyperbolas and parabolas, have been successfully employed by [López-Fernández et al. \(2008\)](#). Now, the hereditary integral can be evaluated by solving the $2N + 1$ ordinary differential equations in Eq. 6.12 at each node s_j . Assuming that the convolution kernel $f(t)$ is analytic for $t > 0$ and locally integrable, [López-Fernández et al. \(2008\)](#) have proved that the method computes the convolution integral with spectral accuracy.

The key advantages of this algorithm is that it only requires $\mathcal{O}(n \log n)$ operations and $\mathcal{O}(n)$ memory, compared with the $\mathcal{O}(n^2)$ operations and memory needed by a “brute force” approach for computing convolutions with n steps of history. Another advantage is that, for the FOV-SLS model, we avoid the direct numerical approximation of the Mittag-Leffler function, and we only need to evaluate the Laplace transform of its derivative, which is given by the following simple expression:

$$\mathcal{L} \left\{ E_{\alpha,0} \left(- \left[\frac{t}{\tau_\sigma} \right]^\alpha \right) \right\} = F(s) = \frac{1}{s^\alpha + \frac{1}{\tau_\sigma}} \quad (6.14)$$

We also report that in simulations where we adopted a “brute force” approach for computing Eq. 6.7, and directly evaluated the Mittag-Leffler function using the algorithm by Podlubny, the solver suffered from instabilities triggered by errors in the approximation of the Mittag-Leffler function, with stability being recovered only when the function was approximated to very high accuracy. Since here we are interested in long-time integration of Eq. 5.1, direct approximation of the Mittag-Leffler function for long times and at the required level of accuracy is further problematic as it introduces a significant computational cost.

In the next section we present patient-specific simulation results of the 1D model in Eq. 5.1, coupled to purely elastic (see Eq. 5.3), and more general SLS (see Eq. 5.10) and FOV-SLS (Eq. 6.6) viscoelastic models. Integer- and fractional-order Kelvin-Voigt models are just limiting cases of SLS models and will be omitted from our attention in what follows.

6.3 Results

6.3.1 Simulation setup

All simulations were performed for a large patient-specific network that provides a detailed geometrical representation of the 50 largest arteries in the human brain, see Fig. 6.1. The data was obtained at the Department of Neurosurgery at Children’s Hospital, Boston MA, USA. First, the 3D geometry was reconstructed from high-resolution magnetic resonance images (MRI) using an “in house” developed software package, and then, the 1D domain was extracted from the centerlines of the 3D vasculature. Consequently, the 1D representation considers arteries as straight tapering tubes, preserving the mean diameter and length of each of the 3D segments, while omitting curvature as well as bifurcation and junction branching angle information. The exact dimensions, model parameters and details of the reconstruction process for this network are given by [Grinberg et al. \(2011a\)](#).

Flow is driven from PC-MRI flowrate measurements at the 4 inlets (see Fig. 6.1): internal carotid (ICA) and vertebral (VA) arteries (see [Grinberg et al. \(2011a\)](#)). The 21 distal outlets are coupled to 3-element windkessel models with parameters obtained from (see [Grinberg et al. \(2011a\)](#)). Given the unavailability of pressure data measurements, these parameters are generally hard to estimate and may have a significant effect on the flow characteristics, leading to even incorrect results (see [Reymond et al. \(2009\)](#)). Here, we have observed that the chosen data set leads to low sensitivities and a physiologically correct solution for our case of interest (see [Grinberg et al. \(2011a\)](#)). The purely elastic response of the arterial wall is quantified by the β parameter, the values of which have also been taken from [Grinberg et al. \(2011a\)](#).

Regarding spatial discretization, the number of DG elements is primarily chosen with respect to the vessel's length. Mesh refinement is performed in regions of high flow gradients, with the final choice of the local polynomial order used (typically 3-8) resulting from mesh independence studies that ensure accurate resolution of the flow dynamics. Following [Grinberg et al. \(2011a\)](#), we have a system with 200 degrees of freedom, and the computed solution does not depend on further resolution refinement. The initial conditions for all simulations are $(A, U)_{t=0} = (A_0, 0)$, and convergence to a periodic solution is obtained after two cardiac cycles. To ensure stability we have chosen a very small time-step $\Delta t = 6 \cdot 10^{-6} s$, a consequence of the high polynomial order that leads to a strict CFL condition and the viscoelastic dissipation introduced by the wall response. The computation of one cardiac cycle takes 30 seconds for elastic or integer-order viscoelastic models, and 15 minutes for fractional-order models (on a desktop with 8 cores of Intel Xeon E5607@2.27GHz).

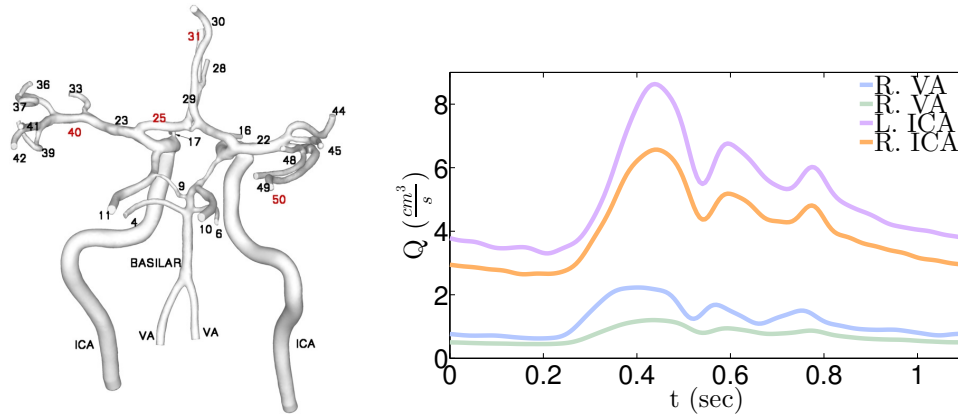


Figure 6.1: *Simulation setup*: Patient-specific cranial network from [Grinberg et al. \(2011a\)](#) and PC-MRI flowrate measurements at the four inlets (L. ICA, R. ICA, L. VA, R. VA)

6.3.2 Comparison of integer- and fractional-order viscoelastic models

We have performed 1D blood flow simulations in the aforementioned cranial network using the integer order SLS, and the fractional order SLS pressure-area relation defined in Eqs. 6.6, 6.7. The integer-order SLS pressure-area relation is determined by the triplet $\{\beta, \tau_\epsilon, \tau_\sigma\}$, while, its fractional-order counterpart, is described by $\{\beta, \alpha, \tau_\epsilon, \tau_\sigma\}$. The nature of these parameters is patient-specific and their values vary among anatomic locations, pathologies, age, etc. (see [Bia et al. \(2005\)](#)). Moreover, their accurate estimation in the clinical setting for various locations in an arterial network is a very challenging task. This leads to a very scarce selection of measured data in the literature, mostly corresponding to the aorta, the carotids and femoral arteries under *ex-vivo* or *in-vivo* conditions.

In this study, we have chosen four different relaxation time parameter sets from the literature, two for the integer order SLS and two for the fractional-order SLS model. The first set (model SLS1) corresponds to the estimations of [Valdez-Jasso et al. \(2011\)](#), where an integer-order SLS model was calibrated to fit measured data for the thoracic descending aorta and the carotid artery under *ex vivo* and *in vivo* conditions in ovine and human arteries. The second set (model SLS2) employs the parameters reported by [Lundkvist et al. \(1996\)](#) for a healthy human femoral artery under *ex vivo* conditions. The third set (model FO-SLS1) is based on the measurements of [Doehring et al. \(2005\)](#), where a fractional-order SLS model was calibrated to *ex vivo* data for porcine aortic valve cusps. Finally, the fourth set (model FO-SLS2) corresponds to the mean values predicted by [Craiem et al. \(2008\)](#) for a fractional-order SLS model fitted to *in-vivo* data from human ascending aorta segments. For comparison purposes we have also included the results obtained by [Grinberg et al. \(2011a\)](#) using the 1D model with a purely elastic pressure-area relation (model Elastic). Since our goal here is the comparison of different viscoelastic models, we have kept the same elasticity parameters reported by [Grinberg et al. \(2011a\)](#) for

all cases considered. Table 6.1 summarizes the viscoelastic models we have employed in this section and their corresponding parameters.

Model	$\tau_\epsilon(\text{sec})$	$\tau_\sigma(\text{sec})$	$\frac{\tau_\epsilon}{\tau_\sigma}$	α
SLS1 (see Valdez-Jasso et al. (2011))	0.050	0.025	2.00	-
SLS2 (see Lundkvist et al. (1996))	29.3	16.9	1.49	-
FO-SLS1 (see Doebling et al. (2005))	1.84	0.076	24.12	0.29
FO-SLS2 (see Craiem et al. (2008))	11.74	7.61	1.54	0.20

Table 6.1: *Comparison of integer- and fractional-order viscoelastic models: Viscoelastic model parameters.*

In Figs. 6.2,6.3 we present the computed waveforms for flowrate and pressure as well as the pressure-area hysteresis loops due to viscoelastic dissipation at representative locations of the arterial network. We observe that all wall models except SLS2 lead to very similar flowrate wave propagation. However, with each model introducing a different amount of viscoelastic dissipation (see Fig. 6.3), the resulting local pressure waves present a variability up to 15%, although all models predict a very similar pressure drop between the inlets and the outlets. This is somewhat expected since the wall response is predominately elastic for all cases. The integer-order viscoelastic models correspond to a fractional order of $\alpha = 1.0$, yet a strong viscoelastic behavior is not observed due to the chosen relaxation times $\tau_\epsilon \leq 2\tau_\sigma$ in both SLS1 and SLS2. Similarly, the relatively low values of the fractional order α in the FO-SLS1 and FO-SLS2 models leads to a dominant elastic response. Note, however, that the change in relaxation times from model SLS1 to SLS2 results to a significant change in the local pressure wave in each vessel (see Fig. 6.3), with the SLS2 model producing a stiffer response and predicting much smaller cross sectional wall displacements. This finding highlights the parametric sensitivity of the integer order SLS model: the discrete relaxation times $\{\tau_\epsilon, \tau_\sigma\}$ dictate the viscoelastic wall

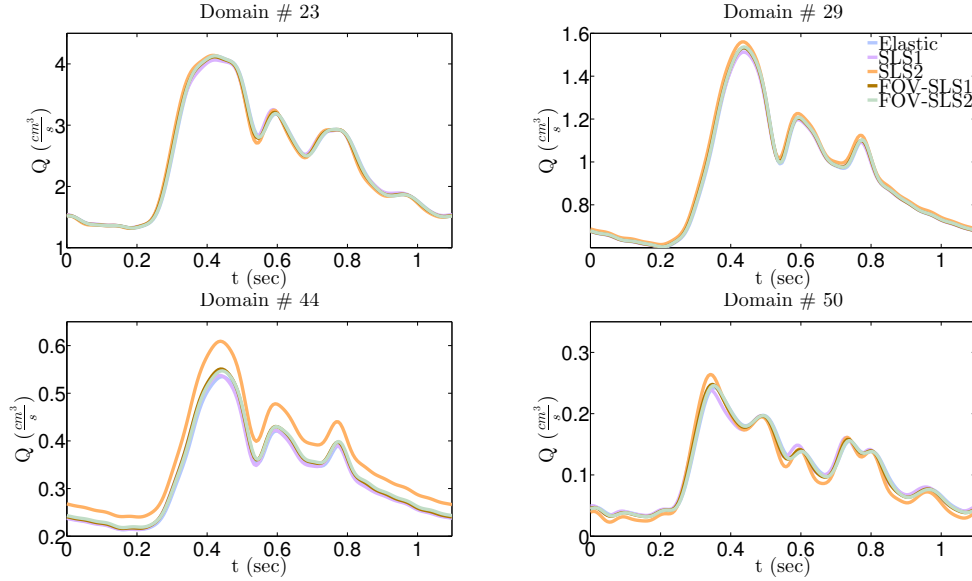


Figure 6.2: *Comparison of integer- and fractional-order viscoelastic models:* Comparison of four viscoelastic model parameter sets: Variability in the computed flowrate at outlets # 23, 30, 44 and 50.

response and strongly depend on the anatomic location. This dependence on the relaxation times introduces a caveat for using the integer order SLS model in simulations of large patient-specific arterial networks for which we only hope to have an estimate of τ_ϵ and τ_σ at few limited anatomic locations.

On the other hand, the behavior of the fractional models is dictated by the fractional order α which controls the interplay between elastic energy storage and viscoelastic dissipation. To better understand the role of α we consider a FO-SLS model and vary the fractional order between $0 \leq \alpha \leq 1$ while keeping the same relaxation timescales with the SLS1 model: $\tau_\epsilon = 0.050\text{s}$ and $\tau_\sigma = 0.025\text{s}$. In Fig. 6.4 we present the computed pressure wave at the left internal carotid artery (L. ICA) and the resulting pressure-area hysteresis loops for different values of α along with the purely elastic and integer order viscoelastic response. As expected, for $\alpha = 0$ we recover the purely elastic stress-strain response, while $\alpha = 1$ returns the solution of the integer order SLS1 model. For $\alpha \leq 0.5$ the response is predominately elastic

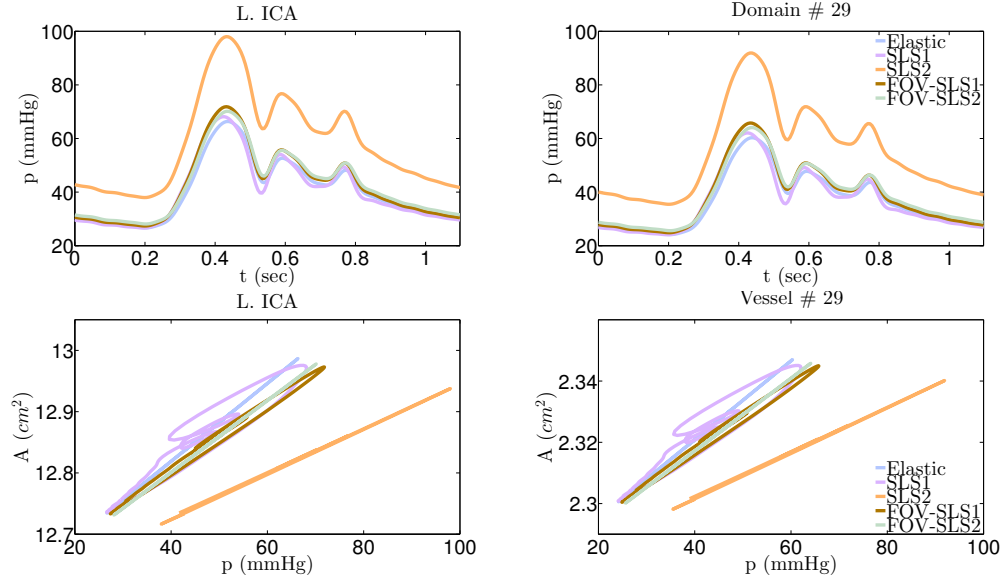


Figure 6.3: *Comparison of integer- and fractional-order viscoelastic models*: Comparison of four viscoelastic model parameter sets: Variability in the computed pressure wave and corresponding pressure-area hysteresis loop at the inlet (L. ICA) and distal outlet # 29.

and the solution is insensitive to the choice of relaxation times, with viscoelastic dissipation being only observed for large relaxation times ratio (see Fig. 6.3 for model FO-SLS1). Increasing the fractional order, the viscoelastic response becomes more important and the choice of relaxation times has a significant impact on the computed solution, both in terms of the pressure waveform and the amount of dissipation introduced in the system. This is illustrated in Fig. 6.4, where we have included results for the FO-SLS model with a larger relaxation time ratio $\frac{\tau_\epsilon}{\tau_\sigma} = 4$, while keeping the same short relaxation time $\tau_\sigma = 0.025s$. This results to a pronounced viscoelastic response that noticeably affects the pressure wave propagation and the pressure-area hysteresis loops.

For $\alpha = 1.0$ the fractional-order SLS model reproduces the integer-order SLS model behavior and suffers from the same sensitivities. However, experimental studies of Craiem and Armentano (2007), Doebling et al. (2005), Craiem et al. (2008) suggest that $\alpha \approx 0.1 \div 0.3$ for larger systemic arteries, while they relate higher

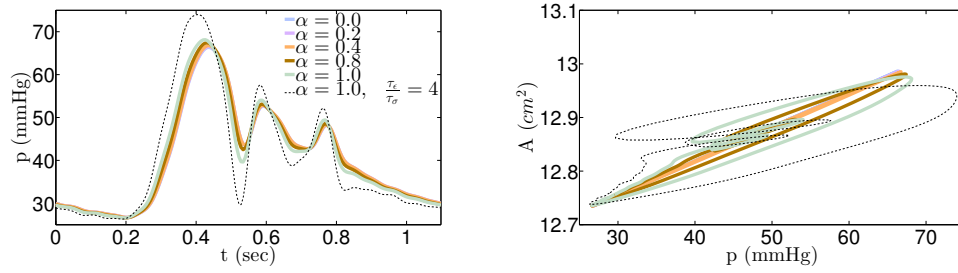


Figure 6.4: *Comparison of integer- and fractional-order viscoelastic models:* Computed pressure and pressure-area hysteresis loop at the inlet (L. ICA) for different values of the fractional order $0 \leq \alpha \leq 1$. The dashed line corresponds to the integer-order model SLS1 (recovered for $\alpha = 1.0$) with a wider relaxation spectrum: $\tau_\sigma = 0.025s$, $\frac{\tau_\epsilon}{\tau_\sigma} = 4$.

fractional orders to smooth muscle activation, which is a key flow auto-regulation agent in arteriolar networks. This may well indicate that muscular small arteries and arterioles exhibit higher fractional orders. Our simulations indicate that for $\alpha \in [0.0, 0.8]$, the effect of the FO-SLS model on the computed blood flow is primarily determined by the fractional order α with low sensitivity on the relaxation times ratio used.

6.3.3 Global sensitivity analysis of the stochastic fractional SLS model

We consider a stochastic fractional order SLS model in which α , $\frac{\tau_\epsilon}{\tau_\sigma}$, and τ_σ are uniformly distributed random variables with value range taken from the literature: $\alpha \in [0, 1]$, $\frac{\tau_\epsilon}{\tau_\sigma} \in [1, 2]$, and $\tau_\sigma \in [0.02, 20.0]$. The corresponding three-dimensional parametric space is discretized with the probabilistic collocation method (PCM) on a tensor product grid generated by the nodes and weights of Legendre polynomials (see [Xiu and Karniadakis \(2002\)](#)). The number of collocation points per random dimension is 8, leading to a total of 512 sampling points. The simulation ran in

parallel on 2048 cores of BG/Q, resulting to a total runtime of 1 hour for 3 cardiac cycles.

In Figs. 6.5,6.6 we present the computed mean flowrate and pressure wave ± 1 standard deviation for two representative locations in the network. The flowrate wave shows a deviation of less than 5% from the mean, indicating that wall viscoelasticity has a minor effect on its propagation. This fact is also in agreement with the results obtained when comparing the integer and fractional-order SLS models (Fig. 6.2). On the contrary, the pressure waveform displays significant variations primarily with respect to the fractional order α , but also with respect to the relaxation times ratio when $\alpha \rightarrow 1$ and the viscoelastic behavior is pronounced. This identifies the role of viscoelastic response as an agent that regulates pressure wave propagation by introducing viscoelastic energy dissipation. This effect is expected to be more pronounced in distal locations where arteries are more muscular and the associated fractional orders are high.

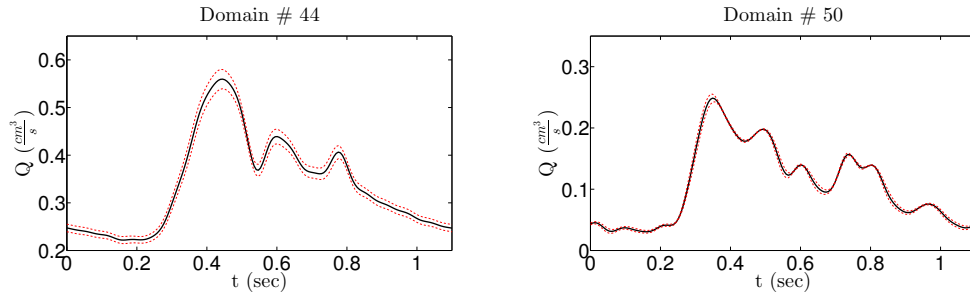


Figure 6.5: *Global sensitivity analysis of the stochastic fractional SLS model*: Computed mean flowrate (in *black*) and a variation of ± 1 standard deviation (in *red*) at distal outlets # 44 and 50.

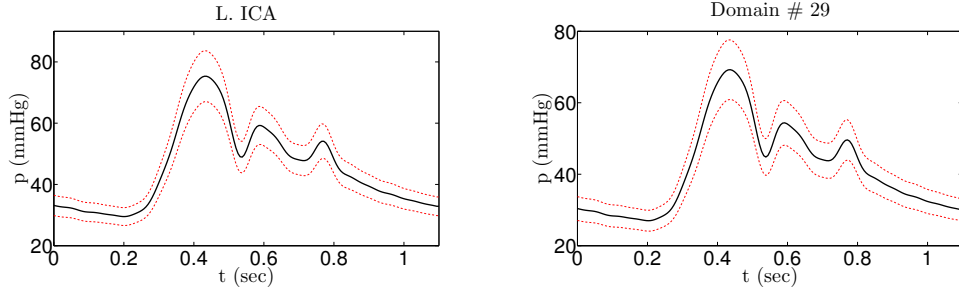


Figure 6.6: *Global sensitivity analysis of the stochastic fractional SLS model*: Computed mean pressure (in *black*) and a variation of ± 1 standard deviation (in *red*) at the inlet (L. ICA) and distal outlet # 50.

To further investigate the parametric sensitivity of pressure wave propagation on viscoelasticity we have performed a global stochastic sensitivity analysis study by computing the standard ANOVA decomposition (see [Yang et al. \(2012\)](#)) of the time-depended solution. Figure 6.7 shows the contribution of each random parameter and their second-order interactions to the total variance of the solution corresponding to the pressure at one of the inlets (left internal carotid artery, L. ICA). Clearly, the fractional order α is the most important parameter, with maximal influence during the systolic phase. This confirms the results presented in the previous sections, illustrating the effect of viscoelastic dissipation (introduced by increasing the fractional order) on the local pressure wave. The second most important parameter is the relaxation times ratio $\frac{\tau_e}{\tau_\sigma}$. In a similar fashion, a larger ratio introduces viscoelastic dissipation which, in turn, affects the local pressure, especially during the systolic phase. The opposite situation is observed for the short relaxation time τ_σ , as it only appears to have a noticeable variance contribution during the diastole. Finally, Fig. 6.7 shows the variance contributions of pairwise interactions of the random parameters, indicating that their effect on the flow is less significant, especially during the systolic phase. We note that although these results are expected for the larger cranial arteries considered here, they provide a clear map of how arterial wall viscoelasticity affects hemodynamics.

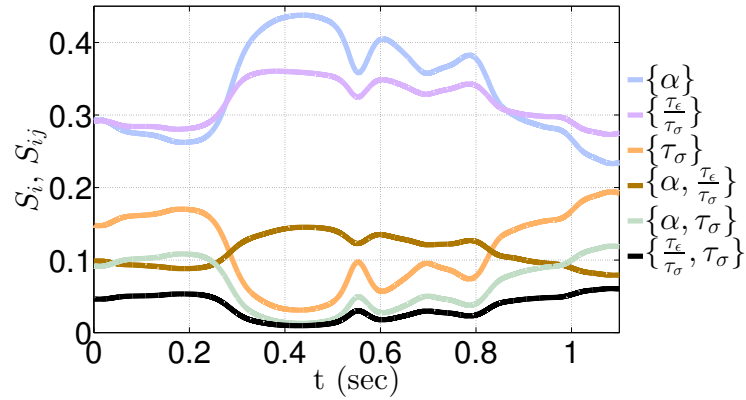


Figure 6.7: *Global sensitivity analysis of the stochastic fractional SLS model*: Variance decomposition of the computed pressure waveform at the inlet (L. ICA). S_i and S_{ij} are the global sensitivity indices of each random parameter and their pairwise interactions, respectively.

6.4 Discussion

We have performed deterministic and stochastic 1D blood flow simulations in a large patient-specific cranial network of compliant arteries corresponding to four different viscoelastic parameter sets. To this end, we have developed a fast parallel solver that supports integer and, for the first time, fractional wall models, enabling us to analyze in detail the effect of viscoelasticity on pulse wave propagation. The solver can be made available to the community upon request and it is written in C++, employing the MPI and OpenMP libraries to enable parallel, multithreaded processing.

First, we compared several models with the same elasticity parameters while the viscoelastic parameters are calibrated with data found in the literature. This comparison study indicated that all models lead to a similar flowrate wave propagation but the resulting local pressure waves are highly affected by the amount of viscoelastic dissipation introduced by each model. We observed that the viscoelastic behavior

of integer-order models is strongly dependent on the choice of the relaxation times, while for fractional models, the response is dictated by the fractional-order. This fact makes fractional-order models very flexible as their conservative or dissipative nature is tuned by a single parameter, the fractional order α , while sensitivity of the solution to the relaxation parameters is only observed as $\alpha \rightarrow 1$. Consequently, fractional-order wall models provide a flexible tool, which may not only accurately capture the static response of the arterial wall but also, by dynamically tuning the fractional order α (which can be a function of time-space in fractional PDEs). One very interesting thrust for future work is neurovascular coupling, in which neural activity could be used to dynamically tune a variable-order fractional viscoelastic wall model model, accounting for smooth muscle activation effects that lead to vasodilation or vasoconstriction and provide a pressure auto-regulation mechanism (see [Witthoft and Karniadakis \(2012\)](#)).

To further investigate the behavior of fractional viscoelastic models, a detailed uncertainty quantification study was performed. Using a stochastic wall model, we have computed a broad range of parametric combinations and reported the mean solution as well as the variabilities observed. We confirmed that wall viscoelasticity has only minor effects on flowrate wave propagation, while pressure waves and resulting wall displacements present variability in their phase and magnitude depending on the amount of viscoelastic dissipation introduced by the fractional order. Finally, by performing a global sensitivity analysis study, we have quantified the relative importance of the viscoelastic parameters and their second-order interactions throughout the span of a full cardiac cycle.

One of the limitations of the present work is that the viscoelastic behavior of each arterial wall model is defined by a single global set of viscoelastic parameters for all vessels. This is due the very limited availability of experimental results for most

of the cranial arteries considered here. We believe that tuning the fractional-order locally for every segment will lead to a more concise modeling approach and we hope that future experimental studies will shed more light on how to appropriately choose it for different anatomic locations, pathologies, etc.

Finally, we must underline that validation is definitely the only path for accepting new tissue models, yet it still remains a very challenging task. In cases where local relaxation measurements are available, simple integer-order viscoelastic models can be calibrated to yield as good results as more general models with more relaxation timescales (see [Valdez-Jasso et al. \(2011\)](#)). However, noninvasive parameter estimation becomes very hard for most distal anatomic locations, making these models suffer from high sensitivity to the relaxation timescales. These issues are key to the practical applicability of any model, and the main reason that may lead one to consider fractional-order models, leveraging on their better sensitivity properties and ability to model continuous relaxation at the expense of a single additional parameter, the fractional order α .

CHAPTER SEVEN

A Fractal-tree Closure Model for Simulating Blood Flow in Large Arterial Networks

7.1 Motivation

The field of computational hemodynamics has undergone great growth in the last 20 years, producing more reliable mathematical models, utilizing modern computer architectures for parallel simulations, and allowing for high fidelity simulations in computational domains reconstructed from patient-specific medical imaging data (see [Xiao et al. \(2013\)](#), [Grinberg et al. \(2013\)](#), [Reymond et al. \(2011\)](#)). While simulations have significantly pushed the boundaries of problem size, performing full scale blood flow simulations on the human arterial network remains intractable hence the need for formulating a proper closure problem for simulations in truncated domains. It is a common practice to reconstruct blood vessel network from medical images covering a particular region, where the resulting models capture only relatively large vessels and additional modeling is required to account for hemodynamics in the truncated vasculature. Specifically, in this article we focus on modeling outflow boundary conditions for simulations in arterial networks with multiple outlets.

Generally speaking, there are two options to properly truncate the computational domain a) use flowrate and/or pressure measurements to directly impose patient-specific data at the outlets, i.e., avoid modeling the truncated arterial tree; and b) use low-dimensional models to *simulate* the flow in the truncated domain. Clearly, in the absence of clinically measured data one has to resort to modeling. Ideally, closure models should be based on sound assumptions with respect to the physics of the problem, be as free as possible of parametric sensitivities, and be fast to solve. In the following we review the current closure models and point to their advantages and limitations.

The widely used simplest representation of the sub-pixel vasculature comes from

zero-dimensional (0D) lumped parameter models. This approach consists of additive combinations of electrical elements, such as resistors R , capacitors C and inductors L , that aim to capture the effective resistance and compliance effects of the neglected arterial networks (see [Alastruey et al. \(2008\)](#), [Reymond et al. \(2011\)](#)). The main advantage of this approach is its implementation simplicity and low computational cost, as it only requires the solution of a linear ordinary differential equation at each terminal outlet. The key disadvantage of this method is the inherent difficulty in estimating the 0D model parameters in absence of local flow measurements at the outlets, in combination with the high sensitivity of the flow model output on these parameters. Recent studies have proposed iterative calibration procedures, in which the problem is solved multiple times until a set of parameters that yields a physiological solution is identified (see [Blanco et al. \(2012\)](#), [Melani and Quarteroni \(2013\)](#), [Lombardi \(2014a\)](#)). Despite their evident limitations, 0D models have been extensively used in the literature (see for e.g., [Reymond et al. \(2011\)](#), [Alastruey et al. \(2008\)](#), [Grinberg and Karniadakis \(2008\)](#), [Blanco et al. \(2012\)](#), [Watanabe et al. \(2013\)](#)), and results for accurately calibrated models have been successfully validated against *in-vivo* measurements (see [Reymond et al. \(2011\)](#)). However, it seems that the obtained windkessel model parameters are not uniform for each arterial network and a calibration procedure has to be repeated to accommodate any network modifications.

An alternative closure model, first put forth by [Olufsen \(1999\)](#), considers modeling the downstream vasculature as a self-similar network of compliant arteries in which a linearized one-dimensional (1D) model is used to estimate the pressure outflow boundary condition as a function of the flowrate time history. The structure of the fractal arterial trees is motivated by the findings of [Zamir \(1999\)](#), while the elasticity parameters vary with the vessel's radius by fitting experimental measurements.

The resulting outflow boundary condition is obtained by computing a convolution operator that relates pressure to flow time history at each of the outlets. Although this approach offers a more detailed representation of the downstream dynamics, it is known to be highly sensitive to the cut-off radius of the structured tree (see [Cousins and Gremaud \(2012\)](#)), suggesting the need for a calibration procedure. Some other limitations of this method include the high number of simulation cycles required to reach a periodic flow state, the costly computation of the convolutions in every time-step, as well as the neglect of the Fahraeus-Lindqvist effect in small arteries and arterioles (see [Pries et al. \(1992\)](#)), and the viscoelastic response of the arterial wall. Recent advances have extended the use of the structured tree model to unsteady problems and explored the use of the Laplace transform to speed up the convergence to a periodic flow state (see [Cousins et al. \(2013\)](#)).

In the case where flow measurements at the outlets are present, [Grinberg and Karniadakis \(2008\)](#) have proposed a simple way of constructing a two-element 0D model that is able to reproduce the measured data *in-silico*. According to [Grinberg and Karniadakis \(2008\)](#), the time varying resistance parameters are derived from the measured flowrates, hence no calibration of the 0D model is required, and simulating 5-10% of a cardiac cycle is sufficient for the mass flow distribution to converge to the reference data.

In this work our main goal is to construct a robust model that eliminates parametric sensitivities and is able to provide a physiological closure to patient-specific hemodynamics simulations with a large number of outlets. We simulate the truncated network by constructing fractal networks of small arteries and arterioles with radii in the range $(10 - 500\mu m)$ that bridge the gap between the outlets of a patient-specific domain and the capillary bed. We simulate hemodynamics using a fully nonlinear 1D solver that accounts for non-Newtonian flow effects and wall viscoelas-

ticity. As this task requires us to resolve flow in networks of thousands to millions of arteries, we have scaled up a parallel 1D Discontinuous Galerkin solver to modern multi-core architectures, moving closer towards the capability of producing *near real-time* numerical solutions.

This chapter we extend the nonlinear 1D blood flow formulation presented in Chapter 5 to performing simulations in massive fractal trees. In Sec. 7.2 we outline the nonlinear 1D closure model and highlight key implementation aspects. In Sec. 7.2 we present the simulation setup for all test cases considered. Then, in Sec. 8.3 we present the main findings of this study. In particular, we address the flow sensitivity issue on the terminal cut-off radius and propose a way to eliminate it by introducing an ansatz on the vessel's length-to-radius ratio for small arterioles. Moreover, we study how the predicted hemodynamics is affected by the Fahraeus-Lindqvist effect (see Pries et al. (1992)) in small arteries and arterioles, and by different choices of boundary conditions at the very distal outlets of the fractal domains. Finally, section 7.6 summarizes our key findings and provides motivation on possible future thrusts of research.

7.2 Fractal-tree closure model

Here we build upon the structured tree framework first put forth by Olufsen (1999) and model downstream flow dynamics by structured trees that aim to bridge the geometric scale gap between outlets with diameters of $\mathcal{O}(cm)$ and the small arterioles that enter the capillary network $\mathcal{O}(\mu m)$. This approach is motivated by the findings of Zamir (1999) and Cassot et al. (2010), which indicate that branching patterns of arteries in the meso-vascular regime resemble fractal laws as follows. A parent vessel

of radius r_0 bifurcates into two daughter vessels with $r_1 = \delta r_0$, $r_2 = \beta r_0$, where

$$\delta = (1 + \gamma^{z/2})^{-1/z}, \quad \beta = \delta \sqrt{\gamma}, \quad \gamma = \left(\frac{r_2}{r_1}\right)^2, \quad r_2 \leq r_1 \quad (7.1)$$

Starting from an outlet with radius r_0 , the resulting fractal-tree is uniquely determined by the parametric set of $\{r_{cut}, L/r, \delta, \beta, \gamma, z\}$, where L/r is the length over radius ratio of each vessel and r_{cut} is the terminal cut-off radius of the tree (see Figure 7.1), β, δ are the parent-daughter radius ratios, γ quantifies the asymmetry of the network, and z is a power law exponent that describes the fluid flow distribution from parent to daughter vessels at each bifurcation (see Zamir (1999), Cassot et al. (2010)). Of these parameters, only r_{cut} may be considered as user input, while the rest may be estimated from medical imaging either in a patient-specific setting or based on population samples. Although δ, β, γ and z seem to be independent of location in the arterial tree (see VanBavel and Spaan (1992), Zamir (1999), Cassot et al. (2010)), the length to radius ratio L/r exhibits a local character, varying significantly among different organs and typically taking values within the range $L/r = 10 - 60$ (see Zamir (1999)). The choice of these parameters determines the size of the generated fractal-tree, and, therefore, the total downstream resistance experienced by each outlet.

The elastic property of the arterial wall is estimated using the experimental fit put forth by Olufsen (1999) that relates the Young's modulus times the thickness of the arterial wall with the vessel radius (see Fig. 3 in Olufsen (1999)). This is the exact relation that has been employed in our work for $r > 500\mu m$, where the curve gives a good fit to the observed experimental data. However, for $r < 500\mu m$ the exponential fit becomes singular as the vessel radius approaches 0. This singularity results to unrealistic predictions, as it implies that small arteries become exponentially stiffer with decreasing radius. To overcome this limitation, we have

assumed that vessels with $r < 500\mu m$ have the same elastic modulus, set to the value predicted by Olufsen's law for $r = 500\mu m$, corresponding to a threshold value of $Eh/r_0 = 7.35 \cdot 10^6 g/sec^2/cm$. Although this choice is arbitrary, we believe it is realistic as small arteries and arterioles are not known to exhibit mechanical responses that may differ by up to 2 orders of magnitude, as predicted from Olufsen's fit for $r \ll 500\mu m$.

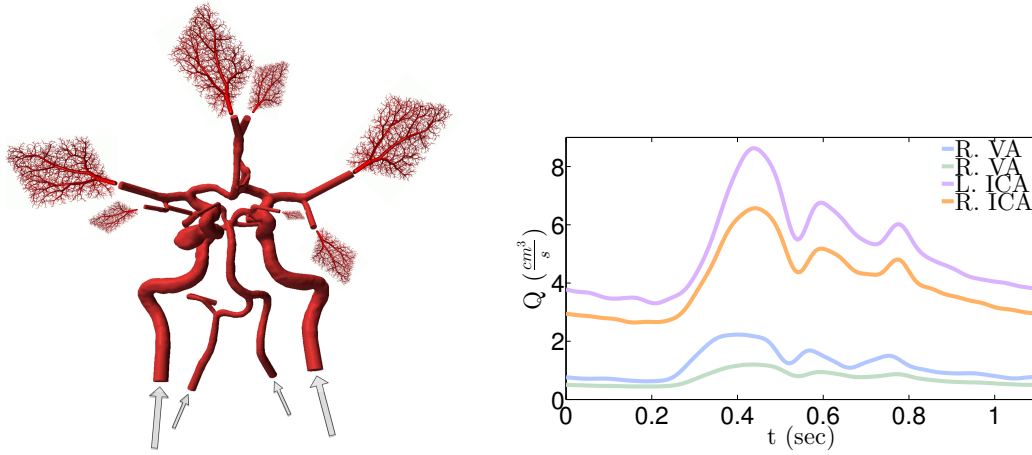


Figure 7.1: *Fractal-tree closure model*: *Left*: Fractal-trees are attached at each outlet of a patient-specific arterial network, resulting to an arterial domain with millions of compliant vessels in which blood flow is modeled using Eq. 5.1. *Right*: Flow is driven by high resolution, patient-specific PC-MRI measurements at the inlets (right vertebral artery R. VA, left vertebral artery L. VA, right internal carotid artery R. ICA and left internal carotid artery L. ICA).

7.3 Implementation aspects

In this work we model flow dynamics in the fractal network using the non-linear conservation law of Eq. 5.1. This allows us to simulate unsteady blood flow in arterial networks spanning multiple spatial scales, to naturally account for the vascular resistance of the arteriolar regime, as well as to resolve the systemic pressure drop, most

of which is known to occur at the level of arterioles (see [McDonald \(1974\)](#)). Moreover, we are able to consider important non-Newtonian flow effects in small arteries and arterioles, as well as to include the effects of wall viscoelasticity. From the implementation perspective, this approach introduces a computational challenge as the resulting arterial network may consist of millions of degrees of freedom, mandating the need for efficient numerical methods and algorithms.

As the cut-off radius is decreased networks consisting of millions of arteries are generated, rendering simulation a non-trivial task. Here, we address this computational challenge by developing a parallel solver based on hybrid programming paradigm that utilizes the MPI and OpenMP libraries, allowing us to decompose the problem among multiple compute nodes as well as explore intra-node parallelism via multithreading. This approach allowed us to scale up our 1D blood flow solver on the IBM Blue Gene/Q architecture and integrate Eq. 5.1 at the rate of $10^{-8}s$ per vessel, per time-step. Based on our experience in simulating 1D flow in networks of thousands to millions arteries, about 100-150 of modern CPUs are adequate for simulating one cardiac cycle in about five minutes. We also envision that further reduction in required CPU time is possible by performing code optimization and scaling to larger number of CPUs.

7.4 Case studies and simulation setup

Flow in the Circle of Willis (CoW): For our first test case we consider a large patient-specific network consisting of the 50 largest arteries in the human brain, see Fig. 7.1. The data was obtained at the Department of [rengachary2005principles](#) at Children’s Hospital, Boston MA, USA. First, the 3D geometry was reconstructed

from high-resolution magnetic resonance images (MRI) using an “in house” developed software package, and then, the 1D domain was extracted from the centerlines of the 3D vasculature. Consequently, the 1D representation preserves the mean diameter and length of each of the 3D segments, while omitting curvature as well as bifurcation and junction branching angle information. The exact dimensions and elasticity parameters of each arterial segment, as well as details of the reconstruction process are given in [Grinberg et al. \(2011a\)](#). The nonlinear 1D blood flow system (Eq. 5.1) is solved in the patient-specific parent network (50 cranial arteries) and the fractal trees attached to each one of the 21 outlets.

The network has 4 inlets (see Fig. 7.1): the two internal carotid (ICA) and two vertebral (VA) arteries, where we apply accurate PC-MRI flowrate measurements (see [Grinberg et al. \(2011a\)](#)). To ensure stability we have chosen a very small time-step $\Delta t = 10^{-6}s$ due to the high polynomial order (ranging between 3-7 depending on the vessel’s length) and the viscoelastic dissipation introduced by the arterial wall response modeled using Eq. 6.6 with $\alpha = 1$ (integer-order Standard Linear Solid model). The elasticity parameters are taken from [Grinberg et al. \(2011a\)](#), the viscoelastic relaxation times were set to $\tau_\epsilon = 0.050s$ and $\tau_\sigma = 0.025s$ (see [Perdikaris and Karniadakis \(2014\)](#)), and the external pressure p_{ext} was taken equal to 15mmHg, corresponding to the mean pressure of the cerebrospinal fluid (see [Reymond et al. \(2011\)](#)).

Flow in a detailed network of the arm: For our second test case we consider a detailed arterial representation of the arm based on the work of [Watanabe et al. \(2013\)](#). This network consists of 246 arteries with 105 terminal outlets, 124 bifurcation points, 18 merging junctions and 85 perforator vessels (see [Watanabe et al. \(2013\)](#), and Fig. 7.2). Each arterial segment is discretized in space using 1 DG element per centimeter of vessel’s length and a polynomial approximation of

second order, while the time-step was taken equal to $\Delta t = 10^{-5}s$. Following Watanabe et. al. we have employed a purely elastic constitutive law parametrized by the data reported by [Watanabe et al. \(2013\)](#). Flow is driven by imposing a physiological flowrate waveform at the inlet of the axillary artery based on information available in the medical literature (see [Watanabe et al. \(2013\)](#)). Similarly, the 1D flow equations (Eq. 5.1) are solved treating the parent arm network (276 arteries) and the fractal trees attached to each one of the 103 outlets, as a whole.

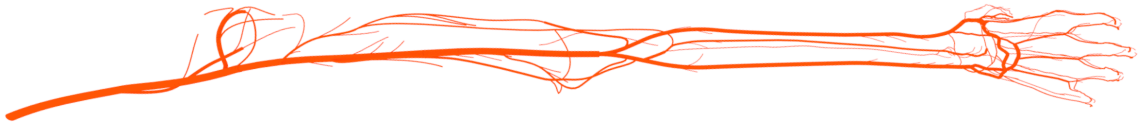


Figure 7.2: *Case studies and simulation setup*: A detailed arterial network of the arm, courtesy of [Watanabe et al. \(2013\)](#) (276 arteries with 105 terminal outlets, 124 bifurcation points, 18 merging junctions and 85 perforator vessels).

In both the arm and cranial cases we have imposed absorbing wave outflow boundary conditions at all distal outlets of the fractal-trees by setting the corresponding Riemann invariant to zero. Blood viscosity is variable in space with the vessel's radius according to the experimental fits of [Pries et al. \(1992\)](#), while the blood hematocrit is considered constant and equal to 0.45. The initial conditions are $(A, U, p)_{t=0} = (A_0, 0, p_{ext})$, and discretization parameters are chosen such that the computed solution does not depend on further resolution refinement.

Starting from each outlet of the of the parent CoW or arm network and given a user specified cut-off radius, we construct several generations of *asymmetrically* bifurcating fractal-trees. The trees are parametrized by $\{r_{cut}, L/r = 50, \delta = 0.9, \beta = 0.6, \gamma = 0.41, z = 2.76\}$, following the values suggested by [Zamir \(1999\)](#). In Table 7.1 we report the size of the resulting arterial network as well as the average number of generations attached to the parent domain for both the CoW and arm cases.

	CoW		Arm	
$r_{cut}(\mu m)$	# of vessels	Avg. # of gen.	# of vessels	Avg. # of gen.
200	1,286	14	-	-
100	8,456	21	1,400	14
50	56,484	27	9,046	19
30	226,522	32	36,228	24
20	682,450	36	109,567	28
10	4,392,484	43	702,188	34

Table 7.1: *Fractal tree closure models*: Total number of vessels and average number of generations attached to each outlet of the parent CoW and arm networks as a function of the fractal-tree cut-off radius r_{cut} .

7.5 Results

In this section we present the main findings of our study. First, we demonstrate the well known issue of sensitivity on the cut-off radius of the fractal-trees (see [Cousins and Gremaud \(2012\)](#)), and propose an effective way of eliminating it by introducing an ansatz for selecting the length-to-radius ratio of the generated small arteries and arterioles. Second, we study two different blood rheology models in order to address the Fahraeus-Lindqvist effect (see [Pries et al. \(1992\)](#)) in small arteries and arterioles and quantify its effect on hemodynamics. Last, we present results on the effect of imposing different types of outflow boundary conditions at the distal outlets of fractal-trees.

7.5.1 Model sensitivity on the cut-off radius

Results of the first set of simulations with fixed L/r ratio are shown in Figure 7.3, The numerical solution is obtained using fractal-trees parametrized by the values suggested by Zamir (1999). Pressure and flow rate waves are probed at the midpoint of the basilar (CoW case) and radial (arm case) arteries, respectively, demonstrating the high sensitivity on the choice of the cut-off radius. In particular, although the model predicts a consistent flow distribution and pressure drop across the arterial network, the computed pressure wave has a non-physiological magnitude, with the systolic pressure exceeding 200mmHg as the cut-off radius is decreased. This observation is in agreement with the results reported by Olufsen (1999), Steele et al. (2007), and Cousins and Gremaud (2012), Cousins et al. (2013), highlighting the main drawback of structured tree-type outflow boundary conditions. A justification of this inconsistency can be given based on the fact that as we decrease the cut-off radius, we greatly increase the size of the generated fractal-trees, leading to a potential overestimation of the total downstream resistance experienced by the outlets of the parent network.

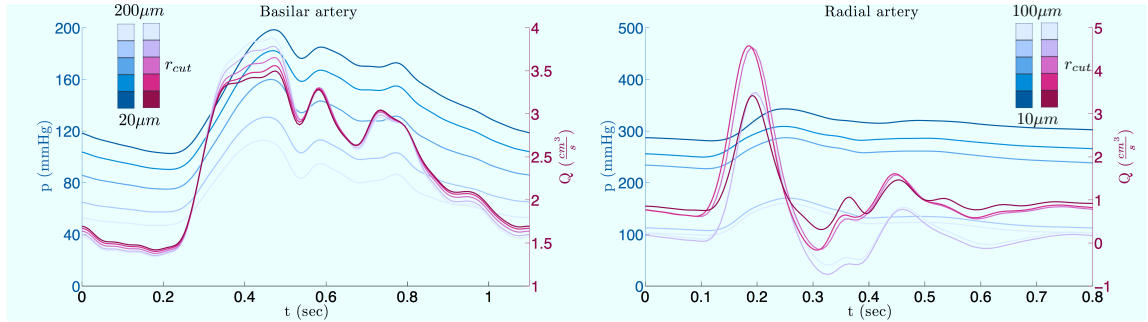


Figure 7.3: *Model sensitivity on the cut-off radius*: Sensitivity on the cut-off radius r_{cut} : Computed pressure (*blue*) and flowrate (*red*) waves for different values of the cut-off radius r_{cut} . *Right*: Basilar artery (CoW case). *Left*: Radial artery (arm case).

7.5.2 Overcoming the cut-off radius sensitivity

One possible solution is to terminate the fractal-tree once the total volume of the generated network reaches a target value that is based on measuring (or estimating) the arterial fraction of the total blood volume considered. This approach can be applied for the CoW case as follows. First, we assume an estimate for the cerebral blood volume of about 120mL (see [Rengachary and Ellenbogen \(2005\)](#)). Then, following the findings of [Ito et al. \(2001\)](#), we take the true arterial fraction of the cerebral blood volume to be about 30%. Using this estimate we can compute the target arterial volume as $120 \cdot 30\% \approx 35\text{mL}$. The main drawback of this approach is that it lacks generality as the suggested termination criterion is only applicable for cases where an estimate of the blood volume and its corresponding arterial fraction at the region of interest is known.

Another possible solution is to adopt a tiered approach for selecting the parameters that govern the structure of fractal-trees, allowing $\{L/r, \delta, \beta, \gamma, z\}$ to vary with the vessel's radius based on experimental data (see [VanBavel and Spaan \(1992\)](#)). This approach was successfully employed by [Steele et al. \(2007\)](#) and [Cousins and Gremaud \(2012\)](#), [Cousins et al. \(2013\)](#), [Cousins \(2013\)](#), although it was reported to be case dependent, potentially leading to inconsistencies if no model calibration is performed (see [Cousins \(2013\)](#)). While our own experience confirms these findings, here we focus our study on testing an alternative way of selecting the length to radius ratio L/r that can effectively eliminate the sensitivity on the cut-off radius and produce physiologically correct results for patient-specific networks without the need of model calibration. To this end, we consider $\{\delta, \beta, \gamma, z\}$ to be constant across the tree, while introducing the ansatz that the L/r ratio is a linearly decreasing function once the radius becomes smaller than a given threshold r^* (see [Figure 7.4](#)).

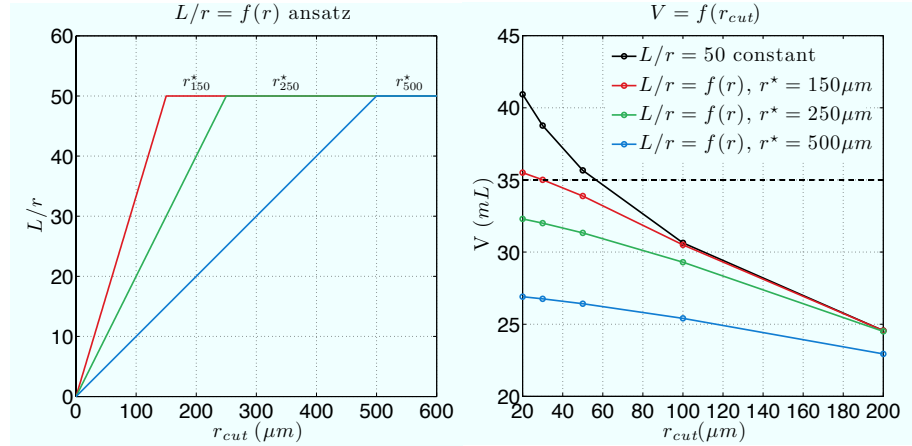


Figure 7.4: *Overcoming the cut-off radius sensitivity: Left:* Decreasing L/r ansatz for different thresholds r^* . *Right:* Volume of the generated arterial tree as a function of the cut-off radius for different L/r relations (CoW case). The dashed line corresponds to the estimated target volume of $35mL$.

In Figure 7.5 we present simulation results where the fractal-trees were generated using the proposed ansatz for the L/r ratio and no prior model calibration was performed. The solution is probed at the same spatial locations, namely the midpoint of the basilar (CoW case) and radial (arm case) arteries, respectively. For the CoW case, we have chosen a threshold of $r^* = 150\mu m$, as this value generates a network that matches the target volume of $35mL$ for a cut-off radius $r_{cut} = 30\mu m$ (see Figure 7.4). We observe that as the cut-off radius is decreased, the propagated flow and pressure waves converge to a physiologically correct plateau. This convergence is best quantified by computing the relative error of the pressure wave between cases with different cut-off radii and a reference case (see table inscribed in Figure 7.5). We also confirm that these results can be also reproduced if we assumed a constant L/r ratio and terminated the fractal-trees once the target volume of $35mL$ is reached ($r_{cut} = 60\mu m$, see Figure 7.4).

For the arm case, there does not exist (to our knowledge) an estimate for neither the blood volume nor for the arterial fraction of the blood volume. Hence, for this

case, no estimate for the target volume can be computed, and r^* needs to be tuned by the user in order for the model to return physiologically correct results. Moreover, pressure wave propagation is expected to be sensitive to the choice of r^* , hence a calibration procedure is needed to obtain physiologically sound results. Although this is a limitation of the proposed framework, we believe that it is not a major one as only a single parameter, r^* , needs to be calibrated. On the contrary, if one wishes to employ the most widely used lumped closure model, the 3-element windkessel, immediately faces the problem of calibrating 206 resistance and capacitance parameters (the parent network consists of 276 arteries and 103 outlets) (see [Watanabe et al. \(2013\)](#), [Reymond et al. \(2011\)](#), [Xiao et al. \(2013\)](#)). Consequently, here we have chosen a L/r relation with a threshold radius of $r^* = 500\mu m$ (see Figure 7.4). Similarly, the flowrate and pressure waves converge to a plateau solution that is no longer sensitive on further decreasing the cut-off radius r_{cut} (see Figure 7.5). Here, although the predicted waveforms are able to reproduce a physiological flow distribution, pressure drop and key wave propagation features such as the dicrotic and the anacrotic notch (see Figure 7.6), the computed pressure exceeds the physiologically correct range. This is because the arm network (courtesy of [Watanabe et al. \(2013\)](#)) is not patient-specific but has been meticulously constructed from anatomical references available in the medical community, and our choice of $r^* = 500\mu m$ was an approximation. Specifically, this network is missing several outlets that need to be considered if we hope to compute a physiologically correct pressure wave without the need of several iterations of a model calibration procedure.

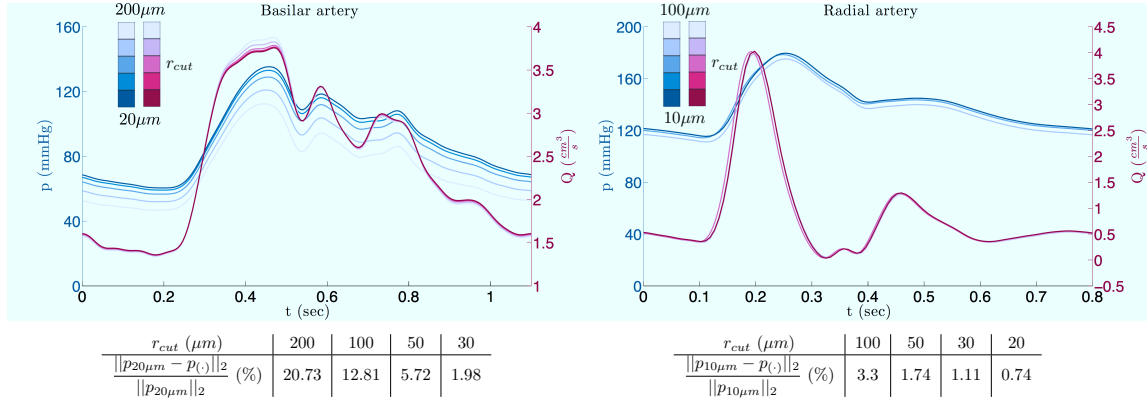


Figure 7.5: *Overcoming the cut-off radius sensitivity*: Waveform convergence with decreasing cut-off radius using the L/r ansatz : Computed pressure (*blue*) and flowrate (*red*) waves for different values of the cut-off radius r_{cut} . *Left*: Basilar artery (CoW case). *Right*: Radial artery (arm case). Inscribed tables contain the relative error ($\mathbb{L}_2$ norm) of the pressure wave as we decrease the cut-off radius (reference cases: $r_{cut} = 20\mu m$ for CoW and $r_{cut} = 10\mu m$ for the arm).

7.5.3 The Fahraeus-Lindqvist effect

In order to accurately model blood flow in small arteries and arterioles ($50\mu m \leq r \leq 500\mu m$), one needs to take into account the dependence of the apparent blood viscosity on the hematocrit and the vessel's radius (the Fahraeus-Lindqvist effect, see [Pries et al. \(1992\)](#)). Here, we have performed two sets of simulations considering the cases of constant blood viscosity (Newtonian flow) versus radius-dependent viscosity (non-Newtonian flow) using the model proposed by [Pries et al. \(1992\)](#). The model is derived from experimental data obtained at high shear rates ($\geq 50s^{-1}$) and provides a clear correlation between blood viscosity in straight tubes, tube diameter and blood hematocrit.

Our findings indicate that the effect of non-Newtonian rheology is primarily manifested in the propagation of the pressure wave. In Figure 7.6 we present computed pressure waves using the L/r ansatz at the left internal carotid artery (CoW case,

$r^* = 150\mu m$) and the axillary artery (arm case, $r^* = 500\mu m$). In both cases the fractal-trees are terminated at $r_{cut} = 20\mu m$, and the solutions have converged to a plateau where they are no longer sensitive to further decreasing the cut-off radius. We observe that modeling blood as a Newtonian fluid has a significant effect on pressure wave propagation, resulting to an effective stiffening of the arterial wall response that leads to an upward shift of the pressure wave, pushing the predicted values away from the physiological range. On the other hand, the radius-depended viscosity model returned a physiologically correct solution, highlighting the fact that capturing non-Newtonian blood behavior is crucial for correctly modeling hemodynamics in small arteries and arterioles. Our results highlight that different rheology models may result to significantly different predictions far upstream in the network due to the cumulative effect of the large arteriolar networks modeled by the fractal tree.

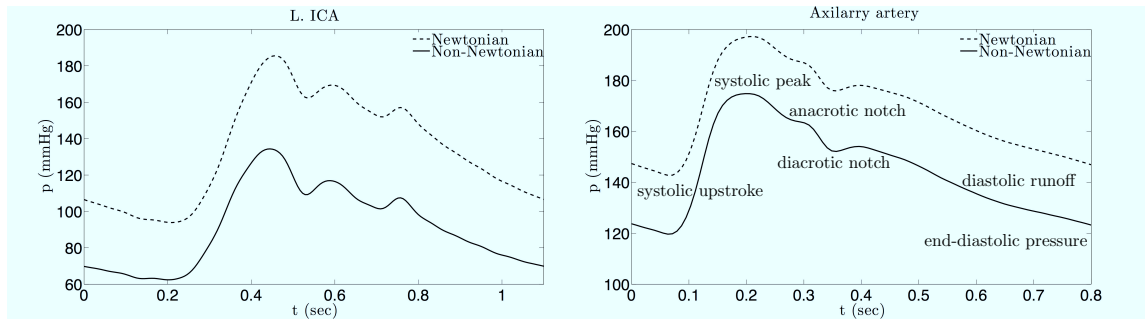


Figure 7.6: *The Fahraeus-Lindqvist effect*: Computed pressure waves for different blood rheology models: radius-depended viscosity suggested by Pries et al. (1992) (non-Newtonian, *solid line*) versus constant blood viscosity (Newtonian, *dashed line*) *Left*: Left internal carotid artery (L. ICA, CoW case). *Right*: Axillary artery (arm case).

7.5.4 Effect of distal outflow boundary conditions

The generated fractal arterial networks typically contain tens of thousands of distal outlets for which we need to provide a proper outflow boundary condition. Here, we

demonstrate that the choice of the prescribed outflow boundary condition has minor effects on the predicted hemodynamics. To this end, we have considered three different types: absorbing wave, constant pressure, and 3-element windkessel boundary conditions. Absorbing conditions are applied by zeroing out the incoming Riemann invariant at the outflow boundary interface, ensuring that no wave reflections will occur. The use of constant pressure boundary conditions is primarily motivated by the fact that the pulsatility of the flow fades out as we approach the capillary bed (arterioles with radius $r < 30\mu m$), and estimates for the pressure values at these locations do exist in the medical literature (see [McDonald \(1974\)](#)). Another possibility is to use simple 3-element windkessel models, consisting of a resistor in parallel with a resistor and capacitor (RCR), with the RCR parameters being roughly estimated based the total resistance of the constructed network and the target pressure values at the outlet (see [Alastruey et al. \(2008\)](#)). Note that due to very small variation of flowrate at terminals of the fractal network the RCR models becomes effectively the resistance model, i.e., the pressure-flowrate relation can be accurately estimated from $P = R_{total}Q$.

In Figure 7.7 we present the computed flowrate and pressure waveforms at representative locations for the CoW and arm networks. In both cases we have used the suggested L/r ansatz (CoW: $r^* = 150\mu m$, arm: $r^* = 500\mu m$), with the fractal-trees being terminated at $r_{cut} = 30\mu m$. The target pressure for the constant pressure and 3-element windkessel boundary conditions was set to 20mmHg, a representative value for small arterioles (see [McDonald \(1974\)](#)). Our results indicate that both the mass flow distribution as well as the local pressure waves converge to a periodic state after two cardiac cycles (starting from zero-flow initial conditions) and show very low sensitivity to the choice of distal outflow boundary conditions. This observation advocates one of the key advantages of the proposed closure model as it seems to

effectively eliminate uncertainties in outflow boundary conditions for hemodynamics.

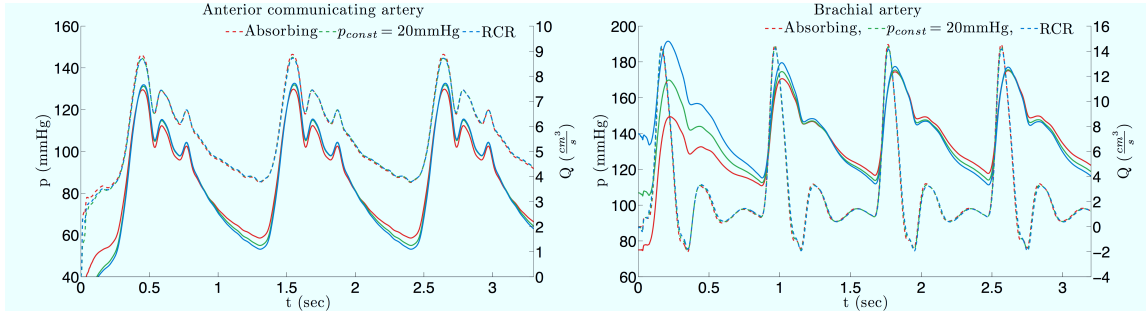


Figure 7.7: *Effect of distal outflow boundary conditions:* Effect of distal outflow boundary conditions: Computed pressure (*solid lines*) and flowrate (*dashed lines*) waves for different types of distal outflow boundary conditions: Absorbing (*red*), constant pressure (*green*), and 3-element RCR windkessels (*blue*). *Left:* Anterior communicating artery (AcoA, CoW case). *Right:* Brachial artery (arm case).

7.5.5 3D-1D coupling in a cerebral network with 10 outlets

We will use the simulation setup first put forth by [Grinberg and Karniadakis \(2008\)](#) in the context of testing a proposed RC lumped parameter model as an outflow boundary condition for arterial networks with multiple outlets. This 3D patient-specific geometry contains 20 cranial arteries reconstructed from high-resolution computed tomography angiography (CTA) images.

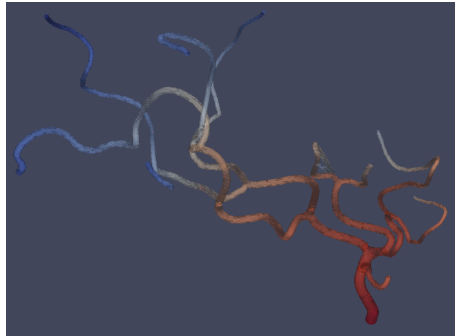


Figure 7.8: *3D-1D coupling in a cerebral network with 10 outlets:* Network of 20 cranial arteries with 1 inlet (red) and 10 outlets.

	3D domain	1D domain
Number of elements	111214	9689
Polynomial order	5	3
Non-dimensional Δt	0.025	0.005
Number of processors	8176	10
Average cpu-time per time-step	0.52s	0.0008s
Wall clock time per one cycle	4.3hrs	-

Table 7.2: *3D-1D coupling in a cerebral network with 10 outlets*: Simulation properties.

The 3D computational domain (Fig.16) is driven by imposing a Womersley flow-rate at its inlet with $Ws = 4.95$, $Re = 200$, and $T = 0.575sec$. Each of the outlets is coupled to a fractal 1D network generated as described in Sec. 7.2 with a cutoff radius of $r_{cut} = 100\mu m$, and a constant $L/r = 50$ ratio. At the distal end of each fractal network we attach a 0-D RCR windkessel model and set its parameters by crudely extrapolating the values published by [Reymond et al. \(2011\)](#) as a function of the outlet radius. The 1D network accounts for 9689 arterial domains which have being discretized in space using 1DG element per domain. Table 7.2 summarizes the properties of this simulation. For comparison purposes we have also considered a purely 3D simulation of the same topology using a 0-D RC model at the outlets (see [Grinberg and Karniadakis \(2008\)](#)).

In Tables 7.3, 7.4 we have summarized the measured errors in the pressure drop Δp between the inlet and each of the 10 outlets, and the measured flow rate at each of the 10 outlets, respectively. Here, the objective is to compare the errors for different values of the cut-off radius of the fractal trees, $r_{cut} = 300, 200, 100\mu m$. In all cases we observe that all modeling approaches return very close predictions for both the measured pressure drops and flowrates. However, we do observe significant variation in the amplitude of the predicted pressure wave. In Figs. 7.9(a),(b) we present the computed pressure wave over one cardiac cycle at each of the outlets for

Δp	$r_{cut} = 200\mu m$			$r_{cut} = 300\mu m$			# of 1D vessels
Outlet	$\mathbb{L}_1$	$\mathbb{L}_2$	$\mathbb{L}_\infty$	$\mathbb{L}_1$	$\mathbb{L}_2$	$\mathbb{L}_\infty$	$100\mu m \rightarrow 200\mu m \rightarrow 300\mu m$
#1	4.47e-02	3.99e-02	3.76e-02	1.06e-01	9.30e-02	8.57e-02	1151 $\rightarrow$ 209 $\rightarrow$ 67
#2	1.75e-02	1.38e-02	-9.76e-05	3.19e-02	2.53e-02	-2.82e-04	2611 $\rightarrow$ 369 $\rightarrow$ 119
#3	6.54e-02	4.39e-02	-8.61e-03	1.15e-01	7.77e-02	-1.50e-02	649 $\rightarrow$ 117 $\rightarrow$ 39
#4	8.92e-03	5.95e-03	5.64e-03	2.13e-02	1.39e-02	1.17e-02	389 $\rightarrow$ 67 $\rightarrow$ 21
#5	5.37e-04	1.22e-02	9.58e-03	2.93e-03	2.60e-02	2.20e-02	1511 $\rightarrow$ 255 $\rightarrow$ 87
#6	1.14e-03	9.74e-03	8.80e-03	5.69e-03	2.08e-02	1.96e-02	1319 $\rightarrow$ 211 $\rightarrow$ 67
#7	1.05e-01	8.04e-02	-2.44e-05	1.91e-01	1.48e-01	1.31e-03	1319 $\rightarrow$ 211 $\rightarrow$ 67
#8	1.87e-01	1.28e-01	-1.98e-02	6.67e-01	4.33e-01	-8.13e-02	529 $\rightarrow$ 87 $\rightarrow$ 29
#9	9.31e-06	1.14e-02	8.16e-03	2.01e-03	2.418e-02	1.74e-02	211 $\rightarrow$ 37 $\rightarrow$ 11
#10	3.05e-03	8.27e-03	4.23e-03	1.42e-02	2.01e-02	6.58e-03	1319 $\rightarrow$ 211 $\rightarrow$ 67

Table 7.3: *3D-1D coupling in a cerebral network with 10 outlets*: Relative errors of the pressure gradient Δp between the inlet and each of the 10 outlets of the 3D domain. We consider $r_{cut} = 100\mu m$ to be our reference case.

Q	$r_{cut} = 200\mu m$			$r_{cut} = 300\mu m$			# of 1D vessels
Outlet	$\mathbb{L}_1$	$\mathbb{L}_2$	$\mathbb{L}_\infty$	$\mathbb{L}_1$	$\mathbb{L}_2$	$\mathbb{L}_\infty$	$100\mu m \rightarrow 200\mu m \rightarrow 300\mu m$
#1	4.21e-02	4.20e-02	4.07e-02	4.21e-02	4.20e-02	4.07e-02	1151 $\rightarrow$ 209 $\rightarrow$ 67
#2	4.26e-02	4.12e-02	-1.39e-02	4.26e-02	4.12e-02	-1.39e-02	2611 $\rightarrow$ 369 $\rightarrow$ 119
#3	1.79e-01	1.74e-01	-5.71e-02	1.79e-01	1.74e-01	-5.71e-02	649 $\rightarrow$ 117 $\rightarrow$ 39
#4	1.08e-02	1.05e-02	8.55e-03	1.08e-02	1.05e-02	8.55e-03	389 $\rightarrow$ 67 $\rightarrow$ 21
#5	8.33e-03	1.29e-02	1.80e-02	8.33e-03	1.29e-02	1.80e-02	1511 $\rightarrow$ 255 $\rightarrow$ 87
#6	8.70e-03	1.16e-02	1.57e-02	8.70e-03	1.16e-02	1.57e-02	1319 $\rightarrow$ 211 $\rightarrow$ 67
#7	1.22e-01	1.20e-01	-3.17e-02	1.22e-01	1.20e-01	-3.17e-02	1319 $\rightarrow$ 211 $\rightarrow$ 67
#8	2.69e-01	2.65e-01	-7.46e-02	2.69e-01	2.65e-01	-7.46e-02	529 $\rightarrow$ 87 $\rightarrow$ 29
#9	5.23e-04	7.68e-03	8.78e-03	5.23e-04	7.68e-03	8.78e-03	211 $\rightarrow$ 37 $\rightarrow$ 11
#10	3.79e-03	6.36e-03	2.71e-03	3.79e-03	6.36e-03	2.71e-03	1319 $\rightarrow$ 211 $\rightarrow$ 67

Table 7.4: *3D-1D coupling in a cerebral network with 10 outlets*: Relative errors of the flow-rate measured at each of the 10 outlets of the 3D domain. We consider $r_{cut} = 100\mu m$ to be our reference case.

a full 3D simulation with 0-D RC boundary conditions versus the coupled 3D-1D case with fractal trees, using $r_{cut} = 100\mu m$, and a constant $L/r = 50$ ratio. We observe that the the 3D-0D model returns a pressure field that is higher than what is physiologically expected. This is primarily attributed to the miscalibration of the the R, C parameters at each outlet. On the other hand, the 3D-1D coupled approach using the proposed fractal tree closure model is able to return a physiological pressure field at all outlets. This difference is less pronounced for the flow rate wave, as shown in In Figs. 7.10(a),(b), where both 3D and 3D-1D distribute the flow in a similar fashion.

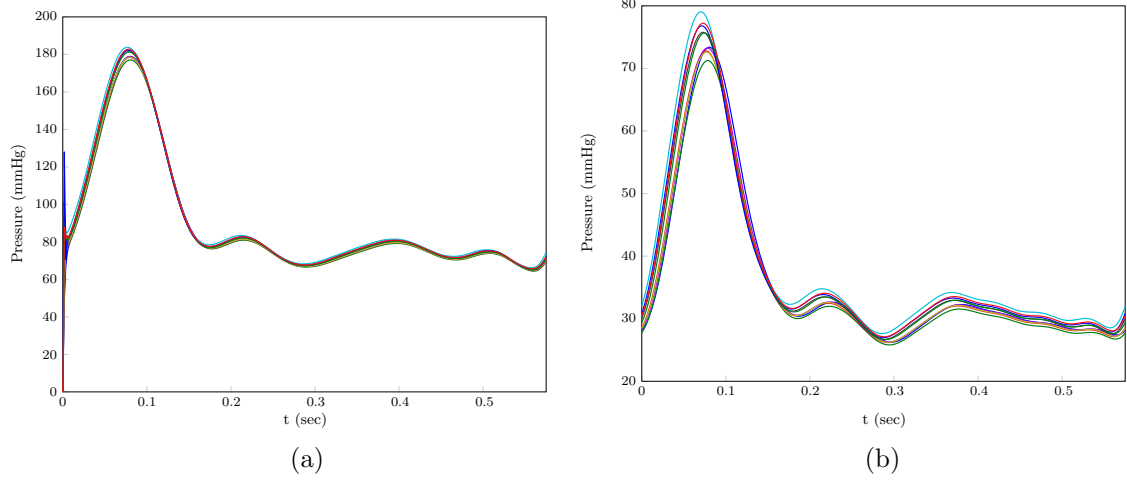


Figure 7.9: *3D-1D coupling in a cerebral network with 10 outlets*: Comparison of computed pressure at the outlets (left: 3D with 0D-RC outflow, right: 3D with 1D elastic fractal networks).

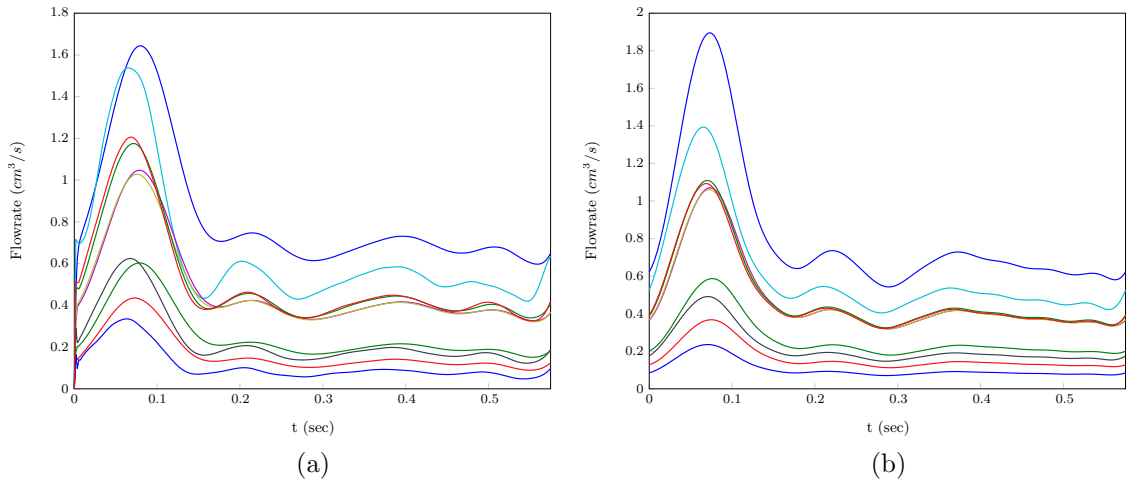


Figure 7.10: *3D-1D coupling in a cerebral network with 10 outlets*: Comparison of computed flow-rates at the outlets (left: 3D with 0D-RC outflow, right: 3D with 1D fractal networks).

7.6 Discussion

We have proposed an outflow closure model for blood flow simulations in large arterial networks, with our main goal being to obtain physiological flow predictions while reducing parametric sensitivities and circumventing costly calibration procedures. The model is based on resolving flow in self-similar networks of millions of small arteries and arterioles that are attached on-the-fly to a given parent vascular network of larger systemic arteries. To our knowledge, this is the first time that nonlinear 1D blood flow simulations have been performed in arterial networks of this size.

While the proposed model minimizes the overall time to accurate solution, it requires facilitation of parallel computers and a scalable solver. At the same time the number of processors (CPU cores) required for reasonable time-to-solution is still very low (less than 1000). This contribution is vital as, not only it allowed simulating trees with hundreds of thousand of arteries in the present work, but it opens the path to addressing many interesting open questions in computational hemodynamics. For instance, the proposed computational framework may be combined with flow and pressure measurements towards solving inverse problems (such as estimating material properties of the arterial wall, estimating the total volume of an arterial tree, etc).

We have suggested a way of addressing cut-off radius sensitivities by introducing a radius-dependent length-to-radius ratio for small arterioles, controlled by a threshold parameter, r^* . This ansatz led to flow predictions that gradually converge to a plateau solution as the cutoff radius is decreased, overcoming cutoff radius sensitivities previously reported for the structured tree outflow boundary condition (see [Cousins and Gremaud \(2012\)](#)).

Our proposed fractal closure model was thoroughly tested for two realistic cases

of blood flow in the Circle of Willis and a detailed network of the arm vasculature. For the CoW case, where the parent network is patient-specific and the fractal model parameters are set according to a target volume estimate, the proposed model returns physiologically correct results without any user intervention or prior calibration. The key point here is that the volume constraint can be achieved without performing any simulation based optimization (as opposed to tuning to match imposed/measured data), leading to significant savings in computation. Moreover, the volume constraint fits well the spirit of the structured tree as it has a physiological basis and enables the study of cases where having a physiologically correct tree is very important (for e.g. heterogeneous autoregulation).

For the arm case, the parent arm network is not-patient specific and a number of outlets may be missing. Moreover, a target volume estimate cannot be readily extracted from the literature, hence the threshold parameter r^* needs to be tuned. The main advantage of tuning r^* versus merely tuning a constant L/r ratio, is that the resulting model exhibits very low sensitivity on the cut-off radius r_{cut} as well as on the outflow boundary conditions used. The important lesson we learned here is that, in the absence of estimates for arterial blood volume in the truncated network, the convergence of flow and pressure waveforms with respect to the cut-off radius can still be obtained.

One limitation of the proposed method is the potential sensitivity on r^* for cases where a target volume estimate is not available. Although this introduces the need for calibration of r^* , we believe that the proposed framework is still viable as it only requires tuning of a single parameter, in contrast to tuning hundreds of resistance and capacitance parameters needed for the RCR model. Another practical limitation of the suggested methodology stems from the computational cost of solving the nonlinear flow equations in networks with hundreds of thousands of arteries. Such

computations can be currently performed in a reasonable time only by utilizing computer clusters with a few hundreds of modern CPUs.

An immediate goal arising from this study is to apply the developed closure model for three-dimensional simulations of a blood flow in compliant arteries. That requires developing stable interface condition between the 3D and the 1D solvers and efficient coupling of the two parallel solvers. Another future thrust is to leverage on the robustness and scalability of the developed solver to address open problems in biomedicine, such as constructing global circulation models, performing high dimensional uncertainty quantification studies, as well as parameter estimation in inverse problems.

CHAPTER EIGHT

Calibration of Blood Flow Simulations via Multi-fidelity Surrogates

8.1 Motivation

Inverse problems are ubiquitous in science. Being inherently ill-posed, they require solution paths that often challenge the limits of our understanding, as reflected by our modeling and computing capabilities. Unlike forward problems, in inverse problems we have to numerically solve the principal equations (i.e., the forward problem) multiple times, often hundreds of times. The complexity in repeatedly solving the forward problem is further amplified in the presence of nonlinearity, high-dimensional input spaces, and massive data-sets; all common features in realistic physical and biological systems. The natural setting for model inversion finds itself within the principles of Bayesian statistics, which provides a formal ground for parameter estimation, i.e., the process of passing from prior belief to a posterior predictive inference scheme in view of data. Here, we leverage the methodology presented in [Chapter 4](#) to design a framework for parameter estimation in blood flow models of the human circulation. Our approach consists of utilizing all available information sources, e.g., experimental measurements, computer simulations, empirical laws, etc., through a general multi-fidelity information fusion methodology in which surrogate models are trained on available data, therefore enabling one to explore the interplay between all such sources.

In general, we model the response of a system as a function $y = f(\mathbf{x})$ of d input parameters $\mathbf{x} \in \mathbb{R}^d$. The goal of model inversion is to identify the parametric configuration in $\mathbf{x}$ that matches a target response y^* . This translates into solving the following optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^d} \|f(\mathbf{x}) - y^*\|, \quad (8.1)$$

in some suitable norm. In practice, $\mathbf{x}$ is often a high-dimensional vector and f is a complex, nonlinear, and expensive to compute map that represents the system's evolving dynamics. These factors render the solution of the optimization problem very challenging and motivate the use of surrogate models as a remedy for obtaining inexpensive samples of f at unobserved locations. To this end, a surrogate model acts as an intermediate agent that is trained on available realizations of f , and then is able to perform accurate predictions for the response at a new set of inputs. Ever since the seminal work of [Sacks et al. \(1989\)](#), the use of surrogates in the design and analysis of computer experiments has undergone great growth, establishing a data-driven mindset for design, optimization, and, as presented in Chapter 3, uncertainty quantification problems. Of particular importance to our work is the approach of [Kennedy and O'Hagan \(2000\)](#) that introduced the use of stochastic auto-regressive maps for building surrogates from a multitude of information sources of variable fidelity. We reckon that this framework enables the meaningful integration of seemingly disparate methodologies, and provides a universal platform in which experiments, simulations, and expert opinion can coexist in tandem. The main challenges here arise from scaling this surrogate-based approach to constructing response surfaces in high-dimensional input spaces and performing tractable machine learning on massive data-sets. Here, our aim is to exploit the developments presented in Chapter 4, towards addressing these challenges and enabling the construction of multi-fidelity surrogates for realistic high-dimensional cases.

In-silico modeling of blood flow in the human vasculature has received great attention over the last 20 years, resulting in the development of computational tools that helped elucidate key biophysical aspects but also aim to provide a customized, patient-specific tool for prediction, intervention, and treatment. Despite great growth in computing power and algorithmic sophistication, the applicability of

such models is limited to truncated arterial domains, as the complexity introduced from considering circulation in the entire arterial network remains intractable (see [Formaggia et al. \(2010\)](#)). Addressing this complexity often leads to the introduction of a series of assumptions, parameters, and simplified models. Consequently, the physiological relevance of our computations directly relies on the calibration of such parameters and models, which, due to the inherent difficulty of them being determined in the clinical setting, remains very empirical, as the resulting hemodynamic problem could admit an infinite number of solutions (see [Quick et al. \(2001\)](#)). To this end, recent efforts for developing model inversion techniques in hemodynamics have been proposed by [Bertoglio et al. \(2012\)](#), [Blanco et al. \(2012\)](#), [Lassila et al. \(2013\)](#), [Lombardi \(2014b\)](#), [Melani \(2013\)](#). A common theme among these efforts is the utilization of reduced order models that can be sampled extensively and with low computational cost, returning a response f that enables a computationally tractable solution to the optimization problem of Eq. 8.1. Among possible reduced order models candidates, the most widely used are nonlinear 1D fluid-structure interactions (FSI) models, linearized 1D-FSI models, and 0-D lumped parameter models (see [Formaggia et al. \(2010\)](#)).

The goal of this work is to incorporate elements of statistical learning towards building a surrogate-based framework for solving inverse problems in hemodynamics, and beyond. Motivated by methods put forth in Chapters 3, 4, we propose an information fusion framework that can explore cross-correlations between variable fidelity blood flow models (e.g., measurements versus 3D-FSI, versus 1D-FSI, versus 0-D models, etc.), allowing for the efficient construction of high-dimensional response surfaces that guide the pursuit for a solution to the optimization problem of Eq. 8.1, while keeping the number of expensive function evaluations at a minimum. Moreover, we aim to demonstrate that this framework is robust with respect

to model misspecification, resulting for example from inaccurate low fidelity models or noisy measurements. Although our main focus here is on parameter estimation for physiologically correct blood flow simulations, the implications of the proposed methodology are far reaching, and practically applicable to a wide class of inverse problems.

This chapter extends the methods presented in Chapters 3, 4 towards designing a tractable methodology from model inversion via multi-fidelity surrogates. In particular, we use Sec. 8.2 to introduce the Efficient Global Optimization algorithm, and highlight how maximizing the expected improvement of the surrogate predictor can lead to efficient sampling strategies for global optimization. The framework is put at test in Sec. 8.3, where we present results for a pedagogical benchmark problem of calibrating the outflow parametrization of a blood flow simulation in a Y-shaped bifurcation.

8.2 Efficient global optimization

Our primary goal here is to utilize the surrogate models generated by the recursive co-kriging formulation towards identifying the global optimum of the optimization problem defined in Eq. 8.1. The probabilistic structure of the surrogate predictors enables an efficient solution path to this optimization problem by guiding a sampling strategy that balances the trade-off between exploration and exploitation, i.e., the global search to reduce uncertainty versus the local search in regions where the global optimum is likely to reside. One of the most widely used sampling strategies in Bayesian optimization that adopts this mindset is the Efficient Global Optimization (EGO) algorithm proposed by Jones et al. (1998). Given a

predictive kriging/co-kriging distribution, the EGO algorithm selects points in the input space that maximize the *expected improvement* of the predictor. To this end, let $f_{\min}^{(t)} = \min\{y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(N_t)}\}$ be the global minimum of the observed response at the t^{th} inference level. Consequently, the improvement of the Gaussian predictor $Z_t(\mathbf{x})$ upon $f_{\min}$ is defined as $I_t(\mathbf{x}) = \max\{f_{\min}^{(t)} - Z_t(\mathbf{x}), 0\}$. Then, the infill criterion suggested by the EGO algorithm implies sampling at locations in the input space that maximize the expected improvement

$$\mathbb{E}[I_t(\mathbf{x})] = [f_{\min}^{(t)} - \hat{y}_t(\mathbf{x})]\Phi\left(\frac{f_{\min}^{(t)} - \hat{y}_t(\mathbf{x})}{s_t(\mathbf{x})}\right) + s_t\phi\left(\frac{f_{\min}^{(t)} - \hat{y}_t(\mathbf{x})}{s_t(\mathbf{x})}\right), \quad (8.2)$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal cumulative distribution and density function, respectively, and s_t is the square root of the predictor variance at level t (see Eq. 2.6). Consequently, the value of $\mathbb{E}[I_t(\mathbf{x})]$ is large if either the value predicted by $\hat{y}_t(\mathbf{x})$ is smaller than $f_{\min}^{(t)}$ or there is a large amount of uncertainty in the predictor $\hat{y}_t(\mathbf{x})$, hence $s_t^2(\mathbf{x})$ is large.

Once a recursive co-kriging cycle has been completed and the final predictor $\hat{y}_s$ and variance $s_s^2(\mathbf{x})$ at level s are known, the expected improvement can be readily computed by Eq. 8.2. Then, the EGO suggests to re-train the surrogates by augmenting the design sets D_t with a set of points that correspond to locations where the expected improvement is maximized. This procedure iterates until a stopping criterion is met. Due to the potentially high computational cost of evaluating f in Eq. 8.1, it is common to use the maximum number of function evaluations as the stopping criterion or the convergence rate of the objective function. Another termination criterion stems from setting a target value for the expected improvement, allowing the next cycle to be carried out only if the expected improvement is above the imposed threshold.

The EGO algorithm can be summarized in the following implementation steps:

Step 1: Given samples of the response $y_t = f_t(\mathbf{x}_t)$, $t = 1, \dots, s$ at all fidelity levels we employ the recursive co-kriging scheme outlined in Sec. 2.3.2 to construct a family of predictive surrogates $\hat{y}_t(\mathbf{x})$.

Step 2: For each surrogate we compute the set of input configurations that maximize the expected improvement, i.e., we solve the optimization problem

$$\max_{x \in \mathbb{R}^d} \mathbb{E}[I_t(\mathbf{x})]. \quad (8.3)$$

Step 3: We evaluate the response $y_t = f_t(\mathbf{x}_t)$, $t = 1, \dots, s$ at the new points suggested by maximizing the expected improvement at each fidelity level.

Step 4: We train again the multi-fidelity surrogates using the augmented design sets and corresponding data to obtain new predictive schemes that resolve in more detail regions of the response surface where the minimum is more likely to occur.

Step 5: Perform Step 2 and check if the desired termination criterion is met. This could entail checking whether a maximum number of function evaluations is reached, whether the minimum of the objective function has converged, or whether the maximum expected improvement is less than a threshold value. If the chosen criterion is not satisfied, we repeat Steps 3-5.

8.3 Results

8.3.1 Y-shaped bifurcation

We consider a benchmark problem for calibrating the outflow parameters in a Y-shaped bifurcation with one inlet and two outlets. The geometry resembles the characteristic size and properties of a carotid bifurcation, yet is kept symmetric to enhance clarity in our presentation. Figure 8.1 presents a schematic representation of the problem setup. In particular, the effect of the neglected downstream vasculature is taken into account through 3-element windkessel models, while the inflow is driven by a physiologically correct flow-rate wave.

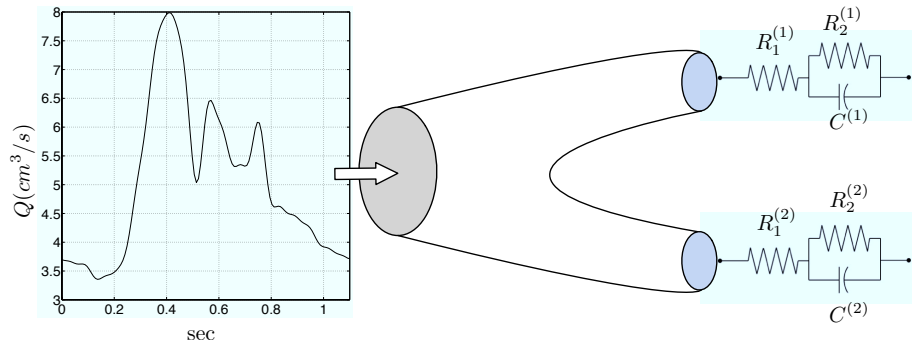


Figure 8.1: *Blood flow in a symmetric Y-shaped bifurcation:* Flow is driven by a physiological flow-rate wave at the inlet, while outflow boundary conditions are imposed through 3-element windkessel models.

Our goal is to calibrate the outflow resistance parameters to obtain a physiologically correct pressure wave at the inlet. To enable a meaningful visualization of the solution process, we confine ourselves to a two-dimensional input space defined by variability in the total resistances imposed at the two outlets, $R_T^{(i)}$, $i = 1, 2$. The total resistance is defined as $R_T^{(i)} = R_1^{(i)} + R_2^{(i)}$, where $R_1^{(i)}$ corresponds to the characteristic impedance of each terminal vessel (see Eq. 5.29). Moreover, the capacitance

parameters are empirically set following the findings of [Grinberg and Karniadakis \(2008\)](#) as $C^{(i)} = 0.18/R_T^{(i)}$, $i = 1, 2$. To this end, our objective is to identify a pair of $R_T^{(i)}$ that matches a physiologically correct systolic pressure of $p_s^* = 126\text{mmHg}$ at the inlet. Assuming no prior knowledge on possible sensible combinations in the input $\mathbf{x} = [R_T^{(1)}, R_T^{(2)}]^T$, we assume a large input space that spans four orders of magnitude: $\mathcal{X} = [10^6, 10^{10}]^2$ measured in $(Pa \cdot s/m^3)$, and consider the following optimization problem

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmin}} |p_s^* - p_s(\mathbf{x})|^2. \quad (8.4)$$

First, we employ a nonlinear 1D-FSI model in order to construct a detailed representation of the response surface of the error in systolic pressure, i.e., $f(\mathbf{x}) = |p_s^* - p_s(\mathbf{x})|^2$, that will be later used as a reference solution to asses the accuracy and convergence of the proposed multi-fidelity model inversion techniques. Here, the choice of studying hemodynamics using 1D models is motivated by their ability to accurately reflect the interplay between the parametrization of the outflow and the systolic pressure at the inlet, yet at a very low computational cost (see [Sherwin et al. \(2003\)](#), [Perdikaris et al. \(2014\)](#)). To this end, Figure 8.2 shows the resulting response surface obtained by probing an accurate nonlinear 1D-FSI solver on 10,000 uniformly spaced samples of the input variables. Since we have considered a symmetric geometry, the resulting response surface exhibits symmetry with respect to the diagonal plane, and can be visibly subdivided using the path of its local extrema into 4 different subregions. The first part is confined in a flat square region near the origin, suggesting that the solution is relatively insensitive to the choice of inputs, as any combination in the inputs within this region results to a reasonably small error in the systolic pressure. The second and third parts correspond to the

symmetric rectangular regions defined by further increasing one of the two resistance parameters. There, the response starts to exhibit sensitivity to the increasing input parameter, leading to a noticeable error in the systolic pressure. Lastly, in the fourth region, the solution exhibits strong sensitivity on the inputs, leading to an explosive growth of the systolic pressure deviation as the resistance parameters take larger values. Interestingly, a closer look at the path of local extrema reveals a subtle undershoot in the response surface along the interface of the aforementioned subregions. Although any input combination within the first subregion results in a very small systolic pressure error, the global solution $\mathbf{x}^*$ of the minimization problem in Eq. 8.4 resides right at the interface of the four subregions (see Fig. 8.2). Note that this topology of the global response is likely to pose serious challenges to any gradient-based optimization strategy, leading to a potentially very large number of function evaluations until convergence, especially in the absence of a very accurate initial guess.

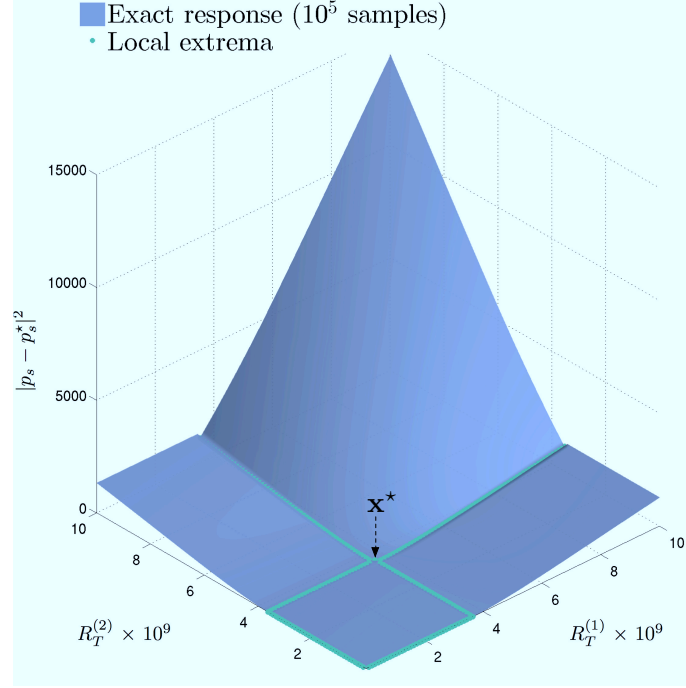


Figure 8.2: *Blood flow in a symmetric Y-shaped bifurcation*: Detailed response surface and identification of the minimum error in the inlet systolic pressure as a function of the total resistance parameters imposed at the outlets. The surface is constructed by probing a nonlinear 1D-FSI model on 10,000 uniformly spaced samples in the space of inputs $\mathcal{X}$.

Now, we turn our attention to solving the optimization problem introduced by Eq. 8.4 using the proposed surrogate-based framework. To this end, in order to build multi-fidelity surrogates for the systolic pressure $y = p_s(\mathbf{x})$ we have considered probing two models: high-fidelity solutions are obtained through the aforementioned nonlinear 1D-FSI solver, while the lower fidelity response is measured through a linearized 1D-FSI model. Here, we emphasize that the linearized model has been purposely derived around a biased reference state, returning erroneous predictions that deviate from the correct solution up to 30%. This is to demonstrate that the proposed methodology is robust with respect to misspecification in the lower fidelity observations.

Initially, we perform 20 linearized low-fidelity simulations, supplemented by 5 high-fidelity realizations, all randomly sampled within the bounds that define the space of inputs. Then, based on the computed response $y_t(\mathbf{x}_t)$, we train a two-level surrogate using the recursive co-kriging scheme (see Sec. 2.3.2) with a stationary Matérn 3/2 auto-correlation function (see Rasmussen (2006)). In order to ensure that any non-stationary effects are filtered out of the data, we use a logarithmic warping of the inputs following the ideas put forth by Snoek et al. (2014). The resulting response surface of the error in the systolic pressure at the inlet is presented in comparison with the reference solution in Fig. 8.3(a). Once the surrogate predictor and variance is available, we can compute the spatial distribution of the corresponding expected improvement using Eq. 8.2. This highlights regions of high expected improvement, as illustrated in Fig. 8.3(b), thus suggesting the optimal sampling locations for the next iteration of the EGO algorithm (see Sec. 8.2). Notice that the maximum expected improvement in this zeroth iteration is already attained very close to the global minimum of the reference solution (see Fig. 8.2).

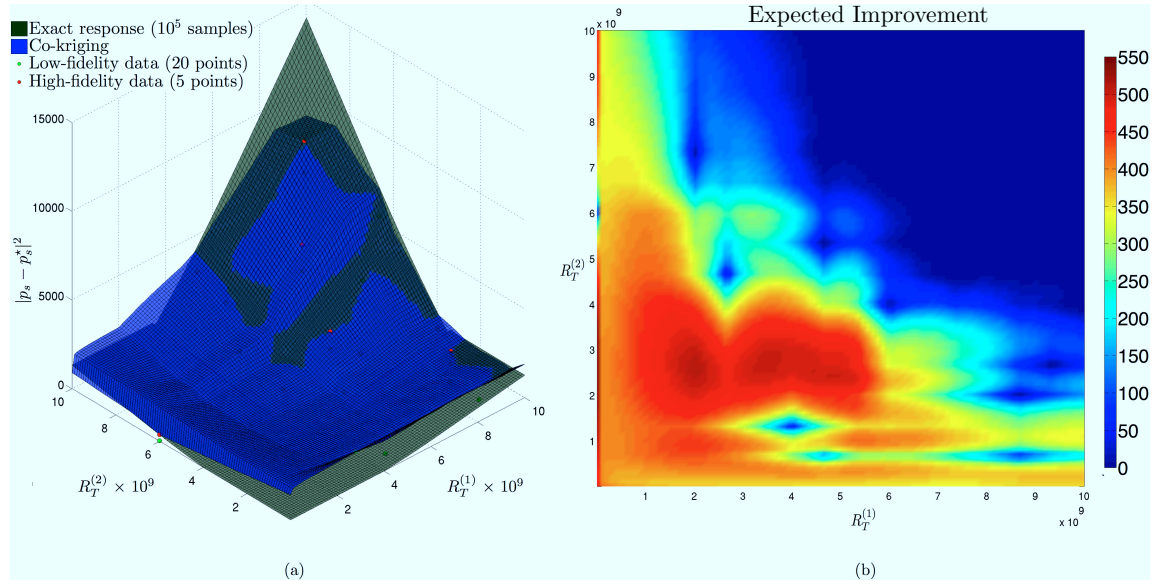


Figure 8.3: *Blood flow in a symmetric Y-shaped bifurcation*: Zeroth iteration of the EGO algorithm. *Left*: Exact solution versus the co-kriging predictor for the inlet systolic pressure error, trained on 100 low-fidelity and 25 high-fidelity observations. *Right*: Map of the corresponding expected improvement.

The next step involves performing an additional set of low- and high-fidelity simulations at the suggested locations of high expected improvement (see Sec. 8.2). The resulting observations are augmented to the existing set of data, and the recursive co-kriging predictor is re-trained to absorb this new information. This procedure is repeated until a termination criterion is met. Here, we have chosen to exit the iteration loop once the minimum of the predicted response surface (i.e., the deviation of the inlet systolic pressure from a target value) is less than $|p_s^* - p_s(\mathbf{x})|^2 < 10^{-3}$.

Figure 8.4(a) presents the resulting response surface of the error in the inlet systolic pressure after 4 iterations of the EGO algorithm. The predictor is trained on 100 low-fidelity and 25 high-fidelity realizations, that clearly target resolving the region of the response surface where the global minimum is likely to occur. Similarly, Figure 8.4(b) presents the spatial distribution of the expected improvement after the new data has been absorbed. By comparing Figs. 8.4(b), 8.3(b) we observe that within 4 iterations of the EGO algorithm the expected improvement has been reduced by a factor greater than 5, effectively narrowing down the search for the global minimum in a small region of the response surface. In fact, 4 iterations of the EGO algorithm were sufficient to meet the termination criterion, and, therefore, identify the set of outflow resistance parameters that returns an inlet systolic pressure matching the target value of $p_s^* = 126\text{mmHg}$. Here, we underline the robustness and efficiency of the proposed scheme, as convergence was achieved using only 25 samples of the high-fidelity nonlinear 1D-FSI model, supplemented by 100 inaccurate low-fidelity samples of the linearized 1D-FSI solver.

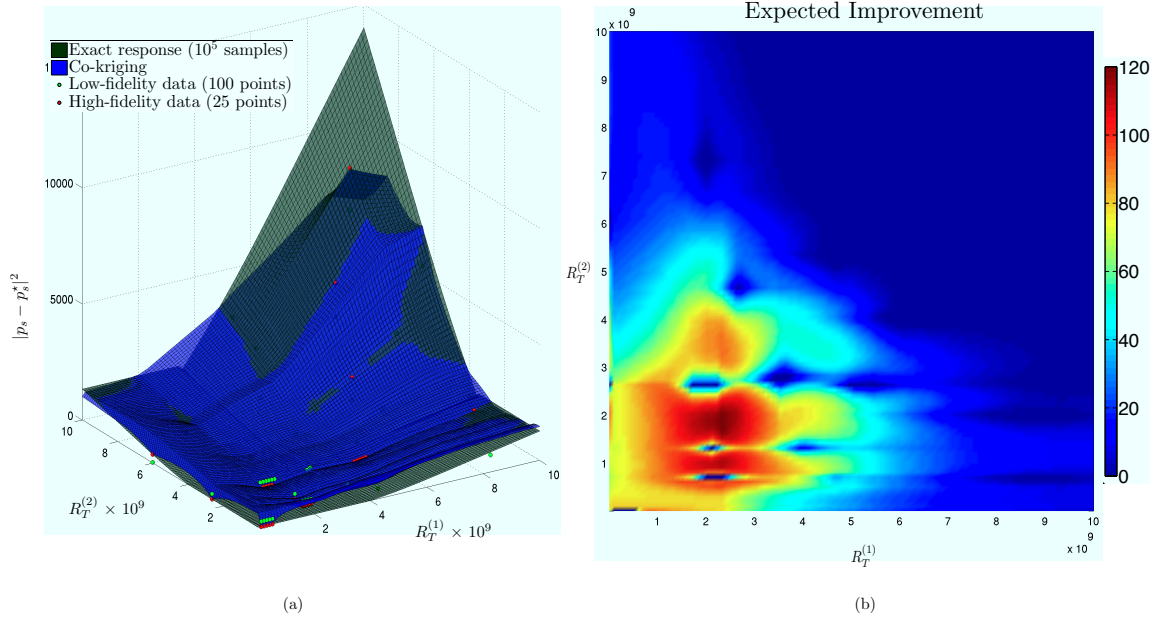


Figure 8.4: *Blood flow in a symmetric Y-shaped bifurcation*: Fourth iteration of the EGO algorithm. *Left*: Exact solution versus the co-kriging predictor for the inlet systolic pressure error, trained on 100 low-fidelity and 25 high-fidelity observations. *Right*: Map of the corresponding expected improvement.

In Table 8.1 we summarize the optimal predicted configuration for the total terminal resistances, along with the relative L_2 -error in the corresponding inlet pressure waveform compared to the optimal reference configuration (see $\mathbf{x}^*$ in Fig. 8.2), over one cardiac cycle, and for every iteration of the EGO algorithm. The convergence of the inlet pressure waveform to the reference solution is demonstrated in Fig. 8.5. As the EGO iterations pursue a match with the target inlet systolic pressure p_s^* , the pressure waveform gradually converges to the reference solution after each iteration of the optimizer.

Iteration No.	$R_T^{(1)} \times 10^9$	$R_T^{(2)} \times 10^9$	$\frac{\ p - p_{\text{opt}}\ _2}{\ p_{\text{opt}}\ _2}$
1	0.001	5.9612	0.1949
2	4.6479	2.6275	0.0945
3	2.0214	2.9275	0.0012
4	3.7183	3.6275	$2.13 \cdot 10^{-4}$

Table 8.1: *Blood flow in a symmetric Y-shaped bifurcation*: Optimal terminal resistances and corresponding relative $\mathbb{L}_2$ -error in the inlet pressure wave over one cardiac cycle, for each iteration of the EGO algorithm. The comparison is done with respect to the inlet pressure wave p_{opt} corresponding to the optimal configuration of resistances predicted by the reference solution obtained using 10,000 high-fidelity samples.

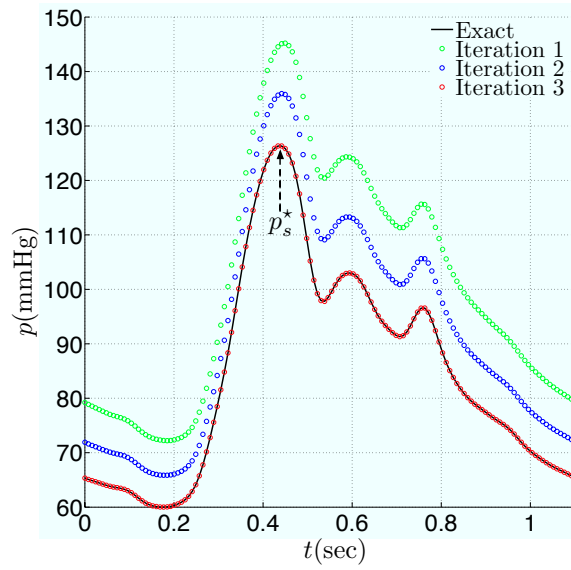


Figure 8.5: *Blood flow in a symmetric Y-shaped bifurcation*: Convergence of the inlet pressure waveform for the first 3 EGO iterations. The exact solution corresponds to the reference results obtained from 10,000 high-fidelity samples, defining the target inlet systolic pressure p_s^* .

8.4 Discussion

We have presented a surrogate-based framework for inverse problems targeting parameter calibration in blood flow simulations. The framework is based on multi-fidelity information fusion and Bayesian optimization algorithms, allowing for the accurate construction of response surfaces through the meaningful integration of variable sources of information, e.g., clinical measurements, *in vitro* experiments, and numerical simulations. Leveraging the properties of the proposed probabilistic inference schemes, we utilize the celebrated Efficient Global Optimization algorithm of [Jones et al. \(1998\)](#) to perform model inversion using only a few samples of expensive high-fidelity model evaluations, supplemented with a number of cheap, but potentially very inaccurate, low-fidelity observations. This approach enables a computationally tractable global optimization framework that is able to efficiently zoom in the response surface in search of the global minimum, while exhibiting robustness to misspecification in the lower fidelity models.

The capabilities of the proposed methodology are demonstrated through a pedagogical benchmark problem involving the parametric calibration of a blood flow solver. To this end, we have considered an symmetric Y-shaped network, consisting of one inlet and two terminal outlet vessels that resemble an idealized carotid bifurcation. The system is closed using a 3-element windkessel model attached to each outlet, and our goal is to identify the configuration of the outflow resistance parameters such that we obtain a physiologically correct pressure field at the inlet. To this end, we have employed two models of different fidelity. The high-fidelity model is an accurate nonlinear 1D-FSI solver, while the low-fidelity model corresponds to an inaccurate linear 1D-FSI solver that has been purposely linearized around a biased reference state. By employing a two-level recursive co-kriging scheme, we are able to

combine the information contained in a few samples of the high-fidelity model, along with a larger set of low-fidelity observations, towards constructing a multi-fidelity surrogate that can estimate the global representation of the error in the inlet systolic pressure as a function of the total resistance parameters in the outflow. This probabilistic representation is then utilized within the EGO loop to iteratively sample the input space towards narrowing down the search for the optimal configuration. To this end, we demonstrated that 4 EGO iterations, requiring 25 high-fidelity and 100 low-fidelity model evaluations, were sufficient to calibrate the outflow parameters, leading to an error in the inlet pressure wave that is less than 10^{-3} when compared against a reference solution obtained using 10,000 high-fidelity observations.

Our exposition here aims to serve as an elucidating example of the key features and capabilities that the proposed framework has to offer. Extending and applying this methodology to more complex and realistic inverse problems requires addressing the challenges of constructing response surfaces in high-dimensions and in the presence of massive data-sets. A possible solution path for addressing these challenges is provided by the methodology presented in [Chapter 4](#).

CHAPTER NINE

Conclusions and Future Work

9.1 Conclusions

In this thesis we have demonstrated how Gaussian process regression and stochastic auto-regressive schemes can provide a launchpad for the development of scalable multi-fidelity information fusion algorithms for modeling the response of complex stochastic dynamical systems. In Chapter 2 we put forth a general framework in which multi-fidelity in physical models and multi-fidelity in probability space can coexist in unison. To this end, in Chapter 3 we proposed fast and accurate predictive algorithms in two and three dimensions based on spectral/ hp element discretizations, and recursive co-kriging schemes that leverage the sparsity introduced by Gaussian Markov random fields. The tractability of the proposed framework was successfully extended to high-dimensional input spaces in Chapter 4 via a data-driven dimensionality reduction technique based on hierarchical functional decompositions. Gearing this approach with $\mathcal{O}(N)$ frequency-domain machine learning algorithms, we were able to develop parallel predictive inference algorithms and demonstrate scaling to extreme cases involving 1,000 input dimensions and $\mathcal{O}(10^5)$ data-sets on a standard desktop computer! These findings set the foundations of a new type of scientific computing which enables a meaningful learning process through exploring and exploiting cross-correlations between a multitude of different methods and information sources, setting a new paradigm in design optimization under uncertainty, data assimilation, model inversion, and beyond.

A parallel thrust aimed at addressing a series of open questions in modeling multi-physics interactions in the human vasculature, focusing on blood flow and arterial biomechanics the brain. To this end, in Chapter 6 we demonstrated how fractional-order models offer an accurate and robust alternative to modeling the viscoelastic response of soft tissue, and we performed a series of systematic computational studies

to quantify the effect of arterial viscoelasticity on blood flow dynamics. Moreover, in Chapter 7 we proposed an effective closure model for simulating blood flow in truncated arterial domains, and scaled up a nonlinear 1D-FSI solver to perform the biggest simulation of this kind ever, involving more than 4 million blood vessels. This methodology, along with the 3D-1D coupling algorithms described in Chapter 5, can provide a stepping stone towards liberating blood flow simulations from the expensive calibration outflow boundary conditions. An alternative approach to parameter estimation was employed in Chapter 7, by addressing the problem of model inversion using multi-fidelity surrogates and Bayesian optimization. This approach leverages directly on the developments presented in Chapters 2-6, and it enables the efficient pursuit to identify global optima while keeping the number of expensive model evaluations at a minimum.

9.2 Future work

We cultivate the vision of building a unified multi-fidelity data-driven framework in high dimensions that can simultaneously pursue the solution of forward and inverse uncertainty quantification problems as well as design and optimization in high-dimensions and under uncertainty. To this end, plan to employ *deep networks* of random fields to perform a seamless fusion of different *information sources* having different levels of fidelity, such as finite elements, particle methods, and noisy experimental data. The objective is to represent the QoI $Y(x)$ in terms of a hierarchy of models connected in an appropriate way. Deep networks can be formally defined as a composition of vector-valued random fields $f^{(i)}(x)$ in $\mathbb{R}^m$ ($i = 1, \dots, L$), with each field being drawn independently from suitable prior distribution. Most of the work that has been done so far refers to deep networks of Gaussian random fields (see

the recent work of [Damianou and Lawrence \(2012\)](#), [Duvenaud et al. \(2014\)](#), [Wilson et al. \(2011\)](#)). However, we can also *estimate* their distribution by using observation models and epi-splines (see [Royset and Wets \(2014b,a\)](#)) approaches, within different layers of the network. The mapping from the state space input variables $x \in \mathbb{R}^n$ (e.g., the design variables) to the output of the network can be represented formally as

$$Y(x) = f^{(L)} \left(f^{(L-1)} \left(\dots f^{(2)} \left(f^{(1)}(x) \right) \dots \right) \right). \quad (9.1)$$

Note that $Y(x)$ is a random field which will allow us to make inferences, e.g., on deterministic quantities such as *risk measures*. The feature map $f^{(i)}$ within each *layer* of the network can be generalized and also become itself a function of the state variables x , i.e., $f^{(i)}(f^{(i-1)}, x)$ (see Fig. 9.1). This allows us to avoid certain pathologies exhibited by deep Gaussian networks. Specifically, the representational capacity of networks in the form (9.1) tends to capture fewer degrees of freedom as the number of layers L increases (see, e.g., [Duvenaud et al. \(2014\)](#)). It is often convenient to set up *observation models* for some or all layers, e.g., in the form

$$y^{(1)}(x) = f^{(1)}(x), \quad y^{(2)}(x) = f^{(2)}(f^{(1)}(x), x), \quad \dots,$$

where $y^{(i)}(x)$ represents data, either deterministic or stochastic. In this way, some or all layers are *not hidden*, i.e., we can inject additional information at each level i for calibration or validation purposes. For example, such observation models can be used in an epi-spline inference process to estimate the (possibly non-Gaussian) distribution of the random field $f^{(i)}$, or more simply to learn its covariance structure - if we assume that $f^{(i)}$ is Gaussian. Thus, the deep network sketched in Fig. 9.1 is really a network of information sources at different levels of fidelity, which are combined together to build an approximation of $Y(x)$. A very particular class of

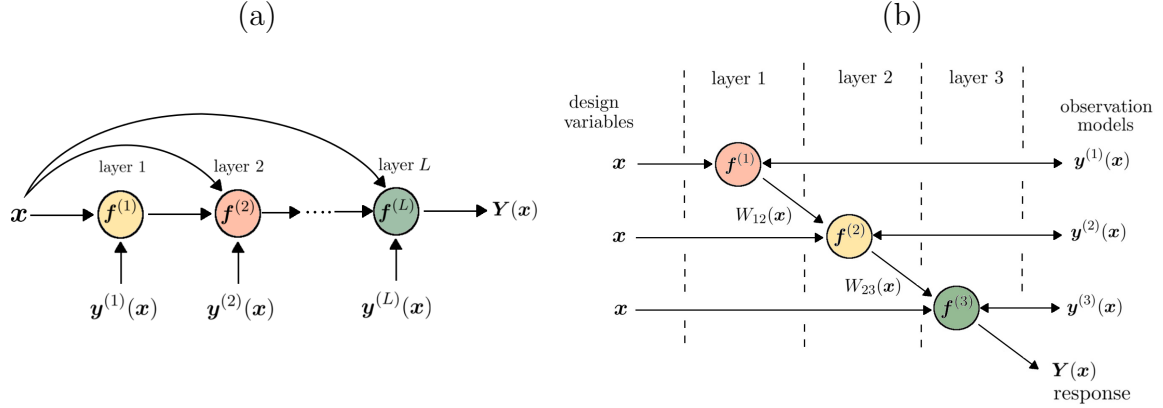


Figure 9.1: *Deep networks*: (a) Sketch of a deep network of vector-valued random fields $f^{(L)} \in \mathbb{R}^m$. In this particular example we have an observation model $y^{(L)} = f^{(L)}$ for each layer, but we can also have situations in which we have hidden (unobserved) layers. The state space variables (e.g., the design variables) are denoted by $x \in \mathbb{R}^n$. The multi-objective response - random field approximating the QoI by $Y(x)$. The deep network allows us to perform information fusion of different information sources by using scalable machine learning algorithms. (b) Recursive co-kriging method as a particular instance of a deep linear Gaussian network.

feature maps $f^{(i)}$ is the linear one, for which we have

$$f^{(1)}(x), \quad f^{(2)}(f^{(1)}, x) = W^{(1)}(x)f^{(1)}(x), \quad f^{(3)}(f^{(2)}, x) = W^{(2)}(x)f^{(2)}(x), \quad \dots \quad (9.2)$$

where $W^{(i)}(x)$ are $m \times m$ matrices with i.i.d. Gaussian entries. In this case, the full input-output map (9.1) takes the form

$$Y(x) = W^{(L-1)}(x) \cdots W^{(1)}(x)f^{(1)}(x), \quad (9.3)$$

and it obviously yields a Gaussian response $Y(x)$. Within the class of networks (9.3) we have, for example, the *recursive co-kriging* of [Le Gratiet and Garnier \(2014\)](#) - a more efficient version of the well-known auto-regressive inference scheme proposed by [Kennedy and O'Hagan \(2000\)](#) that was extensively used throughout this thesis. The recursive co-kriging can be seen as a deep linear network of Gaussian processes in which there is an observation model for each layer (see Fig. 9.1(b)). The network

basically assimilates information, e.g., coming from stochastic numerical codes at different levels of fidelities, within each layer for the purpose of calibrating each (Gaussian) feature map $f^{(j)}$.

It is important to emphasize that the observation models $y^{(i)} = f^{(i)}$ allows us to train $f^{(i)}$ with both *deterministic* and *stochastic* data. In addition, if we have *hidden layers* then the training procedure is applied to the an entire stack of layers. For example, we can simultaneously train layers 1 and 2 (1 is hidden) by using the observation model $y^{(2)} = f^{(2)}(f^{(1)}(x), x)$.

The ultimate goal of this research thrust is to set the stage for the next generation of design and decision making principles in engineering and biomedical applications. An integral part of our future work will be focused on scaling these ideas to realistic scenarios that require the solution of optimization problems involving *thousands of decision variables* in addition to *hundreds of thousands of uncertain parameters*. A key aspect of this goal is to also address inference problems in which the observed outputs are no longer scalar quantities, but could possibly be vectors, or even continuous fields. This is motivated by the recent work of [Nguyen and Peraire \(2015\)](#) which provides a framework for seamlessly integrating measurements and stochastic PDE models towards building accurate inference schemes for functional regression.

This new paradigm for multi-fidelity stochastic modeling allows to go well beyond linear maps, recursive co-kriging methods and Gaussian process regression. This yields new classes of inherently non-linear (non-Gaussian) multi-fidelity information fusion algorithms, which allow us to address surrogate model building in high dimensions. We expect that the implications of such developments will be far reaching and break new ground in uncertainty quantification, design optimization, model inversion, and beyond.

Bibliography

- J. Alastruey, K. Parker, J. Peiró, and S. Sherwin. Lumped parameter outflow models for 1-D blood flow simulations: effect on pulse waves and parameter estimation. *Communications in Computational Physics*, 4(2):317–336, 2008.
- T. Atanacković, S. Konjik, L. Oparnica, and D. Zorica. Thermodynamical restrictions and wave propagation for a class of fractional order viscoelastic rods. In *Abstract and Applied Analysis*, volume 2011. Hindawi Publishing Corporation, 2011.
- H. Baek and G. E. Karniadakis. Sub-iteration leads to accuracy and stability enhancements of semi-implicit schemes for the Navier–Stokes equations. *Journal of Computational Physics*, 230(12):4384–4402, 2011.
- H. Baek and G. E. Karniadakis. A convergence study of a new partitioned fluid–structure interaction algorithm based on fictitious mass and damping. *Journal of Computational Physics*, 231(2):629–652, 2012.
- C. Bertoglio, P. Moireau, and J.-F. Gerbeau. Sequential parameter estimation for fluid–structure problems: Application to hemodynamics. *International Journal for Numerical Methods in Biomedical Engineering*, 28(4):434–455, 2012.
- D. Bia, I. Aguirre, Y. Zócalo, L. Devera, E. C. Fischer, and R. Armentano. Regional differences in viscosity, elasticity, and wall buffering function in systemic arteries: pulse wave analysis of the arterial pressure-diameter relationship. *Revista Española de Cardiología (English Edition)*, 58(2):167–174, 2005.
- K. Binder and D. Heermann. *Monte Carlo simulation in statistical physics: an introduction*. Springer Science & Business Media, 2010.
- P. Blanco, S. Watanabe, and R. Feijóo. Identification of vascular territory resistances in one-dimensional hemodynamics simulations. *J. Biomech.*, 45(12):2066 – 2073, 2012.
- D. Bolin and F. Lindgren. Wavelet Markov models as efficient alternatives to tapering and convolution fields. In *Preprints in Mathematical Sciences*. Citeseer, 2009.
- C. Bron and J. Kerbosch. Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, 1973.

- S. Čanić, C. J. Hartley, D. Rosenstrauch, J. Tambača, G. Guidoboni, and A. Mikelić. Blood flow in compliant arteries: an effective viscoelastic reduced model, numerics, and experimental validation. *Annals of Biomedical Engineering*, 34(4):575–592, 2006.
- F. Cassot, F. Lauwers, S. Lorthois, P. Puwanarajah, V. Cances-Lauwers, and H. Duvernoy. Branching patterns for arterioles and venules of the human cerebral cortex. *Brain res.*, 1313:62–78, 2010.
- W. Cousins. *Boundary Conditions and Uncertainty Quantification for Hemodynamics*. PhD thesis, North Carolina State University, 2013.
- W. Cousins and P. Gremaud. Boundary conditions for hemodynamics: The structured tree revisited. *J. Comput. Phys.*, 231(18):6086–6096, 2012.
- W. Cousins and T. P. Sapsis. Reduced order prediction of rare events in unidirectional nonlinear water waves. *Journal of Fluid Mechanics*, (under review), 2015a.
- W. Cousins and T. P. Sapsis. The unsteady evolution of localized unidirectional deep water wave groups. *Physical review E*, (under review), 2015b.
- W. Cousins, P. Gremaud, and D. Tartakovsky. A new physiological boundary condition for hemodynamics. *SIAM J. Appl. Math.*, 73(3):1203–1223, 2013.
- D. Craiem and R. L. Armentano. A fractional derivative model to describe arterial viscoelasticity. *Biorheology*, 44(4):251–263, 2007.
- D. Craiem, F. J. Rojo, J. M. Atienza, R. L. Armentano, and G. V. Guinea. Fractional-order viscoelasticity applied to describe uniaxial stress relaxation of human arteries. *Physics in medicine and biology*, 53(17):4543, 2008.
- N. A. Cressie and N. A. Cassie. *Statistics for spatial data*, volume 900. Wiley New York, 1993.
- A. Damianou and N. Lawrence. Deep gaussian processes. *arXiv preprint arXiv:1211.0358*, 2012.
- J. De Baar, R. P. Dwight, and H. Bijl. Speeding up Kriging through fast estimation of the hyperparameters in the frequency-domain. *Computers & Geosciences*, 54: 99–106, 2013.
- K. DeVault, P. A. Gremaud, V. Novak, M. S. Olufsen, G. Vernieres, and P. Zhao. Blood flow in the circle of Willis: modeling and calibration. *Multiscale Modeling & Simulation*, 7(2):888–909, 2008.
- C. Dietrich and G. N. Newsam. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing*, 18(4):1088–1107, 1997.
- T. C. Doehring, A. D. Freed, E. O. Carew, and I. Vesely. Fractional order viscoelasticity of the aortic valve cusp: an alternative to quasilinear viscoelasticity. *Journal of biomechanical engineering*, 127(4):700–708, 2005.

- N. Durrande, D. Ginsbourger, O. Roustant, and L. Carraro. Additive covariance kernels for high-dimensional Gaussian process modeling. *arXiv preprint arXiv:1111.6233*, 2011.
- D. Duvenaud, O. Rippel, R. Adams, and Z. Ghahramani. Avoiding pathologies in very deep networks. *arXiv preprint arXiv:1402.5836*, 2014.
- A. C. Eringen. Mechanics of continua. *Huntington, NY, Robert E. Krieger Publishing Co., 1980. 606 p.*, 1, 1980.
- J. Foo and G. E. Karniadakis. Multi-element probabilistic collocation method in high dimensions. *Journal of Computational Physics*, 229(5):1536–1557, 2010.
- J. Foo, X. Wan, and G. E. Karniadakis. The multi-element probabilistic collocation method (ME-PCM): Error analysis and applications. *Journal of Computational Physics*, 227(22):9572–9595, 2008.
- L. Formaggia, J.-F. Gerbeau, F. Nobile, and A. Quarteroni. On the coupling of 3D and 1D Navier–Stokes equations for flow problems in compliant vessels. *Computer Methods in Applied Mechanics and Engineering*, 191(6):561–582, 2001.
- L. Formaggia, A. Quarteroni, and A. Veneziani. *Cardiovascular Mathematics: Modeling and simulation of the circulatory system*, volume 1. Springer Science & Business Media, 2010.
- A. Forrester, A. Sobester, and A. Keane. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008.
- A. I. Forrester, A. Sobester, and A. J. Keane. Multi-fidelity optimization via surrogate modelling. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 463, pages 3251–3269. The Royal Society, 2007.
- Y.-C. Fung. *Biomechanics*. Springer, 1990.
- L. Greengard and J.-Y. Lee. Accelerating the nonuniform fast Fourier transform. *SIAM review*, 46(3):443–454, 2004.
- L. Grinberg and G. E. Karniadakis. Outflow boundary conditions for arterial networks with multiple outlets. *Annals of biomedical engineering*, 36(9):1496–1514, 2008.
- L. Grinberg and G. E. Karniadakis. Extrapolation-based acceleration of iterative solvers: Application to simulation of 3D flows. *Communication in Computational Physics*, 9(3):607–626, 2011.
- L. Grinberg, D. Pekurovsky, S. Sherwin, and G. E. Karniadakis. Parallel performance of the coarse space linear vertex solver and low energy basis preconditioner for spectral/hp elements. *Parallel Computing*, 35(5):284–304, 2009.
- L. Grinberg, E. Cheever, T. Anor, J. R. Madsen, and G. Karniadakis. Modeling blood flow circulation in intracranial arterial networks: a comparative 3D/1D simulation study. *Annals of biomedical engineering*, 39(1):297–309, 2011a.

- L. Grinberg, V. Morozov, D. Fedosov, J. A. Insley, M. E. Papka, K. Kumaran, and G. E. Karniadakis. A new computational paradigm in multiscale simulations: Application to brain blood flow. In *High Performance Computing, Networking, Storage and Analysis (SC), 2011 International Conference for*, pages 1–12. IEEE, 2011b.
- L. Grinberg, D. A. Fedosov, and G. E. Karniadakis. Parallel multiscale simulations of a brain aneurysm. *Journal of computational physics*, 244:131–147, 2013.
- T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani. *The elements of statistical learning*, volume 2. Springer, 2009.
- T. J. Hastie and R. J. Tibshirani. *Generalized additive models*, volume 43. CRC Press, 1990.
- H. Ito, I. Kanno, H. Iida, J. Hatazawa, E. Shimosegawa, H. Tamura, and T. Okudera. Arterial fraction of cerebral blood volume in humans measured by positron emission tomography. *Ann. Nucl. Med.*, 15(2):111–116, 2001.
- D. R. Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of global optimization*, 21(4):345–383, 2001.
- D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.
- G. Karniadakis and S. Sherwin. *Spectral/hp element methods for computational fluid dynamics*. Oxford University Press, 2013.
- M. C. Kennedy and A. O’Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000.
- D. Krige. *A statistical approach to some mine valuation and allied problems on the Witwatersrand: By DG Krige*. PhD thesis, University of the Witwatersrand, 1951.
- N. Kuschel and R. Rackwitz. Two basic problems in reliability-based structural optimization. *Mathematical Methods of Operations Research*, 46(3):309–333, 1997.
- T. Lassila, A. Manzoni, A. Quarteroni, and G. Rozza. A reduced computational and geometrical framework for inverse problems in hemodynamics. *International journal for numerical methods in biomedical engineering*, 29(7):741–776, 2013.
- L. Le Gratiet and J. Garnier. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5), 2014.
- M. Levoy. *Volume rendering using the fourier projection-slice theorem*. Computer Systems Laboratory, Stanford University, 1992.
- G. Li, S.-W. Wang, and H. Rabitz. Practical approaches to construct RS-HDMR component functions. *The Journal of Physical Chemistry A*, 106(37):8721–8733, 2002.

- F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.
- D. Lombardi. Inverse problems in 1D hemodynamics on systemic networks: A sequential approach. *Int. J. Numer. Meth. Biomed. Engng.*, 30(2):160–179, 2014a.
- D. Lombardi. Inverse problems in 1D hemodynamics on systemic networks: A sequential approach. *International journal for numerical methods in biomedical engineering*, 30(2):160–179, 2014b.
- M. López-Fernández, C. Lubich, and A. Schädle. Adaptive, fast, and oblivious convolution in evolution equations with memory. *SIAM Journal on Scientific Computing*, 30(2):1015–1037, 2008.
- C. Lubich and A. Schädle. Fast convolution for nonreflecting boundary conditions. *SIAM Journal on Scientific Computing*, 24(1):161–182, 2002.
- A. Lundkvist, E. Lilleodden, W. Siekhaus, J. Kinney, L. Pruitt, and M. Balooch. Viscoelastic properties of healthy human artery measured in saline solution by AFM-based indentation technique. In *MRS Proceedings*, volume 436, page 353. Cambridge Univ Press, 1996.
- R. L. Magin. *Fractional calculus in bioengineering*, volume 149. Begell House Publishers Redding, 2006.
- F. Mainardi. *Fractional calculus and waves in linear viscoelasticity: an introduction to mathematical models*. World Scientific, 2010.
- G. Matheron. Principles of geostatistics. *Economic geology*, 58(8):1246–1266, 1963.
- K. S. Matthys, J. Alastruey, J. Peiró, A. W. Khir, P. Segers, P. R. Verdonck, K. H. Parker, and S. J. Sherwin. Pulse wave propagation in a model human arterial network: Assessment of 1-D numerical simulations against in vitro measurements. *Journal of biomechanics*, 40(15):3476–3486, 2007.
- D. A. McDonald. Blood flow in arteries. 1974.
- A. Melani. *Adjoint-based parameter estimation in human vascular one dimensional models*. PhD thesis, Italy, 2013.
- A. Melani and A. Quarteroni. *Adjoint-based parameter estimation in human vascular one dimensional models*. PhD thesis, EPFL, 2013.
- M. Mohamad, W. Cousins, P. Perdikaris, and T. P. Sapsis. Probabilistic framework for extreme event quantification: nonlinear water waves and the response of ocean structures to rogue waves. (*in preparation*), 2015.
- T. Muehlenstaedt, O. Roustant, L. Carraro, and S. Kuhnt. Data-driven Kriging models based on FANOVA-decomposition. *Statistics and Computing*, 22(3):723–738, 2012.
- K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

- S. Näsholm and S. Holm. On a fractional Zener elastic wave equation. *Fractional Calculus and Applied Analysis*, 16(1):26–50, 2013.
- N. Nguyen and J. Peraire. Gaussian functional regression for linear partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 2015.
- E. Novak and K. Ritter. High dimensional integration of smooth functions over cubes. *Numer. Math.*, 75(1):79–97, 1996.
- M. S. Olufsen. Structured tree outflow condition for blood flow in larger systemic arteries. *American journal of physiology-Heart and circulatory physiology*, 276(1):H257–H268, 1999.
- G. Papadakis. Coupling 3D and 1D fluid–structure-interaction models for wave propagation in flexible vessels using a finite volume pressure-correction scheme. *Communications in numerical methods in engineering*, 25(5):533–551, 2009.
- T. Passerini, M. de Luca, L. Formaggia, A. Quarteroni, and A. Veneziani. A 3D/1D geometrical multiscale model of cerebral vasculature. *Journal of Engineering Mathematics*, 64(4):319–330, 2009.
- A. T. Patera. A spectral element method for fluid dynamics: laminar flow in a channel expansion. *Journal of computational Physics*, 54(3):468–488, 1984.
- P. Perdikaris and G. E. Karniadakis. Fractional-order viscoelasticity in one-dimensional blood flow models. *Annals of biomedical engineering*, 42(5):1012–1023, 2014.
- P. Perdikaris, L. Grinberg, and G. E. Karniadakis. An effective fractal-tree closure model for simulating blood flow in large arterial networks. *Annals of biomedical engineering*, pages 1–11, 2014.
- M. S. Phadke. *Quality engineering using robust design*. Prentice Hall PTR, 1995.
- I. Podlubny. Calculation of the Mittag-Leffler function with desired accuracy. <http://www.mathworks.com/matlabcentral/fileexchange/8738-mittag-leffler-function>. Accessed: 2012-09-12.
- I. Podlubny. *Fractional differential equations: an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications*, volume 198. Academic press, 1998.
- A. Pries, D. Neuhaus, and P. Gaehtgens. Blood viscosity in tube flow: dependence on diameter and hematocrit. *American Journal of Physiology-Heart and Circulatory Physiology*, 263(6):H1770–H1778, 1992.
- C. M. Quick, W. L. Young, and A. Noordergraaf. Infinite number of solutions to the hemodynamic inverse problem. *American Journal of Physiology-Heart and Circulatory Physiology*, 280(4):H1472–H1479, 2001.
- H. Rabitz, Ö. F. Aliş, J. Shorter, and K. Shim. Efficient input-output model representations. *Computer Physics Communications*, 117(1):11–20, 1999.

- R. Raghu, I. E. Vignon-Clementel, C. A. Figueroa, and C. A. Taylor. Comparative study of viscoelastic arterial wall models in nonlinear one-dimensional finite element simulations of blood flow. *Journal of biomechanical engineering*, 133(8):081003, 2011.
- C. E. Rasmussen. Gaussian processes for machine learning. 2006.
- S. S. Rengachary and R. G. Ellenbogen. *Principles of neurosurgery*. Elsevier Mosby, 2005.
- P. Reymond, F. Merenda, F. Perren, D. Rüfenacht, and N. Stergiopulos. Validation of a one-dimensional model of the systemic arterial tree. *American Journal of Physiology-Heart and Circulatory Physiology*, 297(1):H208, 2009.
- P. Reymond, Y. Bohraus, F. Perren, F. Lazeyras, and N. Stergiopulos. Validation of a patient-specific one-dimensional model of the systemic arterial tree. *American Journal of Physiology-Heart and Circulatory Physiology*, 301(3):H1173–H1182, 2011.
- R. T. Rockafellar and J. O. Royset. Engineering decisions under risk-averseness. (*under review*), 2014.
- J. Royset and R. Wets. Multi-variate epi-splines and evolving function identification problems. pre-print, University of California, Davis, 2014a.
- J. Royset and R. J.-B. Wets. From Data to Assessments and Decisions: Epi-Spline Technology. In A. Newman, editor, *INFORMS Tutorials*. INFORMS, Catonsville, 2014b.
- J. O. Royset. On sample size control in sample average approximations for solving smooth stochastic programs. *Comput. Optim. Appl.*, 55(2):265–309, 2013.
- Y. A. Rozanov. *Markov random fields*. Springer, 1982.
- H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. CRC Press, 2005.
- J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical science*, pages 409–423, 1989.
- P. Seshadri, P. Constantine, G. Iaccarino, and G. Parks. Aggressive design: A density-matching approach for optimization under uncertainty. *arXiv preprint arXiv:1409.7089*, 2014.
- S. Sherwin, V. Franke, J. Peiró, and K. Parker. One-dimensional modelling of a vascular network in space-time variables. *Journal of Engineering Mathematics*, 47(3-4):217–250, 2003.
- S. J. Sherwin and M. Casarin. Low-energy basis preconditioning for elliptic substructured solvers based on unstructured spectral/hp element discretization. *Journal of Computational Physics*, 171(1):394–417, 2001.
- C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77(2):439–471, 1988.

- T. W. Simpson, V. Toropov, V. Balabanov, and F. A. Viana. Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come or not. In *12th AIAA/ISSMO multidisciplinary analysis and optimization conference*, volume 5, pages 10–12, 2008.
- J. Snoek, K. Swersky, R. S. Zemel, and R. P. Adams. Input warping for Bayesian optimization of non-stationary functions. *arXiv preprint arXiv:1402.0929*, 2014.
- B. Steele, M. Olufsen, and C. Taylor. Fractal network model for simulating abdominal and lower extremity blood flow during resting and exercise conditions. *Comput. Methods Biomech. Biomed. Engin.*, 10(1):39–51, 2007.
- B. N. Steele, D. Valdez-Jasso, M. A. Haider, and M. S. Olufsen. Predicting arterial flow and pressure dynamics using a 1D fluid dynamics model with a viscoelastic wall. *SIAM Journal on Applied Mathematics*, 71(4):1123–1143, 2011.
- M. L. Stein, J. Chen, M. Anitescu, et al. Stochastic approximation of score functions for Gaussian processes. *The Annals of Applied Statistics*, 7(2):1162–1191, 2013.
- S. Surjanovic and D. Bingham. Virtual Library of Simulation Experiments: Test Functions and Datasets. Retrieved March 20, 2015, from <http://www.sfu.ca/~ssurjano>.
- D. Valdez-Jasso, D. Bia, Y. Zócalo, R. L. Armentano, M. A. Haider, and M. S. Olufsen. Linear and nonlinear viscoelastic modeling of aorta and carotid pressure–area dynamics under in vivo and ex vivo conditions. *Annals of biomedical engineering*, 39(5):1438–1456, 2011.
- E. VanBavel and J. Spaan. Branching patterns in the porcine coronary arterial tree. Estimation of flow heterogeneity. *Circ. Res.*, 71(5):1200–1212, 1992.
- S. Watanabe, P. Blanco, and R. Feijóo. Mathematical model of blood flow in an anatomically detailed arterial network of the arm. *ESAIM, Math. Model. Numer. Anal.*, 47(04):961–985, 2013.
- P. Whittle. STOCHASTIC-PROCESSES IN SEVERAL DIMENSIONS. *Bulletin of the International Statistical Institute*, 40(2):974–994, 1963.
- A. Wilson, D. Knowles, and Z. Ghahramani. Gaussian process regression networks. *arXiv preprint arXiv:1110.4411*, 2011.
- A. Witthoft and G. E. Karniadakis. A bidirectional model for communication in the neurovascular unit. *Journal of theoretical biology*, 311:80–93, 2012.
- N. Xiao, J. D. Humphrey, and C. A. Figueroa. Multi-scale computational model of three-dimensional hemodynamics within a deformable full-body arterial network. *Journal of computational physics*, 244:22–40, 2013.
- D. Xiu and G. Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.
- X. Yang, M. Choi, G. Lin, and G. Karniadakis. Adaptive ANOVA decomposition of stochastic incompressible and compressible flows. *Journal of Computational Physics*, 231(4):1587–1614, 2012.

- Y. Yu, H. Baek, M. L. Bittencourt, and G. E. Karniadakis. Mixed spectral/hp element formulation for nonlinear elasticity. *Computer Methods in Applied Mechanics and Engineering*, 213:42–57, 2012.
- Y. Yu, H. Baek, and G. E. Karniadakis. Generalized fictitious methods for fluid–structure interactions: analysis and simulations. *Journal of Computational Physics*, 245:317–346, 2013.
- M. Zamir. On fractal properties of arterial trees. *Journal of theoretical biology*, 197(4):517–526, 1999.
- Z. Zhang, M. Choi, and G. E. Karniadakis. Error estimates for the ANOVA method with polynomial chaos interpolation: tensor product functions. *SIAM Journal on Scientific Computing*, 34(2):A1165–A1186, 2012.
- T. Ziehn and A. Tomlin. GUI-HDMR—A software tool for global sensitivity analysis of complex models. *Environmental Modelling & Software*, 24(7):775–785, 2009.